***DETERMINANTS OF LOW BIRTH WEIGHT INFANTS AMONG VARIOUS REGIONS OF ETHIOPIA: A MULTILEVEL LOGISTIC REGRESSION MODEL***

By:

Gedlu Belachew

A thesis submitted to Department of Statistics, School of Graduate Studies, College of Natural Science, Jimma University, in Partial Fulfillment for the Requirement of Masters of Science in Biostatistics.

September 22, 2017
Jimma, Ethiopia

***DETERMINANTS OF LOW BIRTH WEIGHT INFANTS AMONG VARIOUS REGIONS OF ETHIOPIA: A MULTILEVEL LOGISTIC REGRESSION MODEL***

M.Sc. Thesis

By:

Gedlu Belachew

Advisor: Dinberu Seyoum (PhD Scholar)

Co-advisor: Fikadu Zewude (M.Sc.)

September, 2017

Jimma, Ethiopia

**DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES**

**JIMMA UNIVERSITY**

As thesis advisors, we have read and evaluated the thesis prepared By **GEDLU BELACHEW under** our guidance, which is entitled **Determinants of low birth weight infants among various regions of Ethiopia.** We recommend that the thesis fulfills the requirements for the degree of Master of Science.

**Dinberu Seyoum (PhD scholar)** --------------------- ------------------------

Main advisor Signature Date

**Fikadu Zewude (MSc.)** --------------------- -----------------------

Co-advisor Signature Date

As the members of the board of examiners of MSc thesis open defense examination, we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as the degree of Master of Science in Biostatistics

--------------------------------------------------- ------------------------------ --------------------------

Name of chairman Signature Date

--------------------------------------------------- ------------------------------- --------------------------

Name of Main Advisor Signature Date

--------------------------------------------------- ------------------------------------ ----------------------

Name of Co-advisor Signature Date

--------------------------------------------------- ------------------------------------ -------------------

Name of internal Examiner Signature Date

--------------------------------------------------- ------------------------------------ ----------------

Name of External Examiner Signature Date

# STATEMENT OF THE AUTHOR

First, I declare that this thesis is a result of my genuine work and all sources of materials used for writing it have been duly acknowledged. I have submitted this thesis to Jimma University in partial fulfillment for the Degree of Master of Science. The thesis can be deposited in the library of the university to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

**Gedlu Belachew**

**Date:** ____________________

**Signature:** ____________________

**Jimma University, Jimma**

## Acknowledgments

First and for most, I would like to thank, my God the almighty who avail me to this moment.

Second, I have great thanks for CSA and its stuffs whose cooperative, honest, accountable and respective customer service quality and yield this quality data.

Third, I have deep application and thanks for my advisor and brother Dinberu Seyoum whose honest, responsive, building, and intellectual comment and advise.

Finally, I would like to thank my friends, Tariku Fersha for his material support, Gora Alemayhu for his experience share of how to analyze data via MLwin software, and my friend Azmeraw Misganaw for his arrengment of  this work.

# Dedication

This thesis is dedicated to Mr.Geremew Muleta (Assist Prof, Head Department of Statistics, Jimma university), to Jimma university and particularly to Hasselt university (Belgium).

## ABSTRACT

**Background:** Low birth weight is defined as weight of child at birth less than 2500 g measured within 24 hours of birth. It is a public health problem affecting 15%-20% of births worldwide. Low birth weight is the cause of 28%-30% of neonatal deaths. Therefore, this study is conducted to determine the prevalence of Low birth weight and associated factors among newborns delivered in Ethiopia.

**Objective:** The objective of the study is determination of Low birth weight infants variations among various regions of Ethiopia using multilevel logistic regression models.

**Methods:** Data is taken from the 2011 Ethiopian demographic and health survey, which is a nationally representative survey of children in the 0-59 month age groups. Three model families, Empty model, Random intercept model and random slope model will be used for the analysis. MQL-1 and PQL-2 estimation methods are likely to be adequate for producing unbiased estimates compared to other methods.

**Results:** The result showed that 53.2% of children were born with low birth weight. Based on the model adequacy test the random slope binary logistic regression model is found to be the best fitting to the data. The variance of the random component model related to the intercept and sex of child variable are statistically significant.

**Conclusion** This study suggests that sex of child, maternal wealth index and maternal no antenatal visit have been found simultaneously statistically significant. But univariate analysis shows that sex of child, maternal wealth index, maternal place of residence, maternal education level, maternal anemia level, multiple birth, maternal age and maternal weigh have been found statistically significant and are varies across region.


Key words: low birth weight; Null model, random intercept model and Random slope model.

LIST OF ACRONYMS

| | |
|---|---|
| ANC | Antenatal care |
| BMI | Body Mass Index |
| CSA | Central Statistical Agency |
| DHS | Demographic and Health Survey |
| D f | Degrees of freedom |
| EDHS | Ethiopian Demographic and Health Survey |
| GLMs | Generalized Linear Model |
| HLM | Hieratical Linear Model |
| HGLM | Hieratical General Linear Model |
| IGLS | Iterative Generalized Least Squares |
| LBW | Low Birth Weight |
| MDG | Millennium Development Goal |
| MQL | Marginal Quasi Likelihood |
| MCMC | Markov Chain Monte Carlo |
| Ms | Mean square |
| ML | Maximum Likelihood |
| PQL | Penalized Quasi Likelihood |
| PAF | Population Attributable Fraction |
| Ss | Sum of squares |
| UNICEF | United Nations International Children’s Emergency Fund |
| UNSCN | United Nations System Standing Committee on Nutrition |
| WHO | World Health Organization |
| WI | Wealth Index |

List of tables

## List of Tables

## List of Figures

# 1. Introduction

## 1.1 Background of the Study

Low birth weight (LBW) is defined as weight of child at birth less than 2500 g measured within 24 hours of birth (WHO). Level of low birth weight is categorized in to three based on weight. Low birth weight is defined as less than 2500 grams, very low birth weight is less than 1500 g and extremely low birth weight is less than 1000 g (WHO 2011). This practical cut-off for international comparison is based on epidemiological observations that infants weighing less than 2,500 grams are approximately 20 times more likely to die than heavier babies (Kramer, 1998). More common in developing than developed countries, a birth weight below 2,500 grams contributes to a range of poor health outcomes (UNICEF/WHO, 2004).

The incidence of LBW is estimated to be 16% worldwide, 19% in the least developed and developing countries and 7% in the developed countries (UNICEF and WHO, 2004). Globally, more than 20 million infants are born with LBW (UNICEF and WHO, 2004). The largest number of LBW babies is concentrated in two regions of the developing world which are Asia and Africa. Seventy-two percent of LBW infants in developing countries are born in Asia, specifically, in South Asia that accounts for half of the LBW, and 22% are born in Africa. The prevalence of LBW in sub-Saharan Africa ranges between 13% and 15%, with little variation across the region as a whole (UNICEF and WHO 2004). In East Africa the prevalence of LBW is 13.5% (UNICEF and WHO, 2004) and in Ethiopia between 2006 and 2010, UNICEF estimated the prevalence of LBW to be 8%. LBW infants are at risk of 40-fold greater chance of dying in the neonatal period, 50 percent greater chance of serious development problems and 5-10 points decrease in IQ Point. It is also associated with long term disabilities, including visual and hearing impairments, cardiovascular disease and diabetes in later life (UNICEF 2002).Low birth weight is closely associated with fetal and neonatal mortality and morbidity, inhibited growth and cognitive development, and chronic diseases later in life. LBW infant in turn are at higher risk of prenatal death, adult hood stunting and the intergenerational effect of malnutrition continues. Intrauterine growth and development is one of the most vulnerable periods in the human life cycle, which accounts for about 50% of causes of LBW (Ramakrishna 2004).

Many factors affect the duration of gestation and fetal growth, and thus, the birth weight. They relate to the infant, the mother or the physical environment and play an important role in determining the birth weight and the future health of the infant (UNICEF 2004).Birth weight is affected to a great extent by the mother's own fetal growth and her diet from birth to pregnancy, and thus, her body composition at conception. Mothers in deprived socioeconomic conditions frequently have low birth weight infants. In those settings, the infant's low birth weight stems primarily from the mother's poor nutrition and health pregnancy, the high prevalence of specific and non-specific infections, or from pregnancy complications, underpinned by poverty. Also engaging in physically demanding work during pregnancy contributes to poor fetal growth (UNICEF 2004). Low birth weight (LBW) can be caused either by premature delivery (short gestation<37 week) or by fetal growth retardation. Known factors for pre-term delivery and fetal growth retardation which are associated with LBW include low maternal food intake, hard physical work during pregnancy, and illness, especially infections. The studies suggest that cigarette smoking, genetic and environmental factors can cause LBW, short maternal stature, very young age, high parity, close birth spacing is all associated factors (Kramer, 2004). Maternal nutrition status is one of the most important risk factors for LBW (Imdad and Bhutta 2013). Poor maternal nutrition is a known cause of LBW accounting for about 50% of cases of LBW in many developing countries. Those with poor nutrition have higher chance of giving LBW baby (Ramakrishna 2004; Muthayya 2009).

LBW is one of the critical issues in Ethiopia that causes many babies short- term and long term health consequences and tend to have higher mortality and morbidity. DHS Ethiopia /2005/ report shows that the percentage of LBW babies has increased in the past five years from 8 percent in 2000 to 14 percent in 2005. LBW is a reasonable well-defined problem caused by factors that are potentially modifiable and the costs of preventing them are well within rich, even in poor countries like Ethiopia. Therefore, it is very important to determine the risk factor of LBW in various communities in the country in order to come up with feasible intervention strategies to minimize the problem.

Certain data will not be continuous like binary and count data, (in this case binary data), and the corresponding outcome variables are categorical and count responses. Such outcome variables will not be normally distributed rather distributed as binomial, Poisson, gamma etc. In addition, in case of multistage or clustered sampling procedure, response variables will be correlated within individuals in the same clusters. EDHS data is a two stage stratified sampling where infants are the first sampling unit in each household within regions. There may be also having regional variations that is; heterogeneity within regions as well as between regions on birth weight. In a study considering the dichotomous outcome in the multilevel model, McCaul et al. (1992) compare dropping out to personal, social, and labor market experiences. Raudenbush and Bryk (2002) present a two-level HGLM example using ‘repetition’ as a binary outcome and different socio-demographic variables as predictors. Wong and Mason (1985) have considered the hierarchical logistic regression model in relation to model a binary response, using the logit function as an outcome in the level-1 model. These researchers illustrate micro- and macro-level models in which micro observations are embedded within macro observations. However, the interpretive procedures they used in these analyses were analogous to those commonly used in the two-level HGLM, with level-1 (micro) embedded within the level-2 (macro) equations. Research clarifies the advantages of the HGLM over the single-level logistic model. For example, Guo and Zhao (2000) maintain that multilevel modeling not only enables one to decompose the total variance associated with the outcome variable into the parts of each level, but also facilitates reducing cluster bias, produces correct standard errors and correct confidence intervals.

## 1.2 Statement of the problem

In Ethiopia the LBW estimate has risen from 2000 to 2005 from 15.0 % to 20.3% with 1.1% increase per year (UNSCN, 2013). It is therefore imperative to identify risk factors for LBW in various communities in the country in order to come up with feasible intervention strategies to minimize the problem. Identification of maternal risk factors associated with LBW is essential in order to guide program planning, and organizing care for mothers and their newborns. It is expected that identifying those risk factors will enable to reverse the increasing trend of LBW in Ethiopia and thereby it’s immediate and long term consequences. The study will be conducted by directly measuring newborns weight within one hour of delivery in randomly selected health centers and hospitals which is to fill the gap of many studies conducted in Ethiopia that uses maternal subjective evaluation of the babies’ size at birth. In Ethiopia, studies done on determinant factors of LBW are very limited. Moreover, many studies have been done regarding this, but not so much in Ethiopia in recent times. Because in Ethiopia there is limited information on distribution of birth weight. In the context of developing countries where institutional delivery is very low, concentrating only on the children weighed at the health facilities creates some informational gap. Therefore, the current study aims at finding the magnitude and the determinants of low birth weight in Ethiopia based on the EDHS data by taking into consideration various maternal socio-economic and children and maternal demographic factors. Moreover, previous studies on this area in Ethiopia were considered about modeling only the fixed effects of covariates without including the random effects and with no considering sampling structures of data. Most of the studies previously done are simply using only the ordinary logistic regression model. Thus, the little magnitude of this service and lack of appropriateness of the model applied for clustered data have generated interest in assessing determinant factors affecting low birth weight by fitting a statistical model that can explain the data in most meaningful manner. This study, therefore, has tried to fill the gaps in understanding the status of child weight at birth by identifying determinant factors of LBW in Ethiopia and assessing the performance of different models using clustered data from EDHS by addressing the following research questions:

This study has been motivated to address the identified research gaps by answering the following research questions:

I. What are the key factors that affect birth weight at each level between each region?

II. How much of the variation of weight is accounted for regional level and household level?

III. How important is this factor in relation to other socio-economic and demographical factors?

IV. Are traditional (single level) and multilevel approach different in identifying determinants of low birth weight?

## 1.3 Objectives

### 1.3.1 General Objective of the study

The general objective of this study is to identify factors that determine the low birth weight infant in Ethiopia based on EDHS data set using multilevel logistic regression models.

### 1.3.2 Specific objectives of the Study

The specific objectives of this study are:

- ✓ To identify socioeconomic and demographic determinants associated with low birth weight.
- ✓ To identify low birth weight variations due to the random effects at household levels and regional levels.
- ✓ To compare results from traditional (single level) and multilevel approach difference in identifying determinants of low birth weight.
- ✓ To fit adequate statistical model for low birth weight.

## 1.4 Significance of the Study

The study will provide useful information that will inform health facilities prepare special care for LBW neonates, most dies within 24 hours of delivery due to absence of adequate and special care. The findings from this study will provide baseline data for further researches and interventions. Also contribute for ministry of health, regional health bureau, administrators and other concerned organizations and stakeholders to give great emphasis to the problem and take appropriate measures towards the initiation of a suitable nutrition and health promotion programs for pregnant women, which contribute its great share for decreasing the prevalence of LBW neonates. In addition, the study may be useful to other researchers as reference material while conducting further studies on similar problems. Identification of maternal risk factors associated with LBW is also essential in order to guide program planning, and organizing care for mothers and their newborns. It is expected that identifying those risk factors will enable to reverse the increasing trend of LBW in Ethiopia and there by its immediate and long term consequences.

# Chapter Two

## 2. Literature Review

### 2.1. Incidence of Low Birth Weight

The prevalence of LBW greatly varies according to settings and situations. A hospital based cross sectional study conducted on LBW in Perambalur shows LBW prevalence 11.67%. The proportion of LBW babies was more in mothers from rural area (71.43%) than urban area (2.86%). The proportion of LBW was high among mothers who had less antenatal checkups during pregnancy which is 22.2% (Shannon et al. 2008). Other similar study in Utter Pradesh, India on 350 samples shows higher prevalence of LBW as 40%. With 38.5% and 31% among mothers who had inter pregnancy interval less than 2 years and more than 2 years respectively. High prevalence was observed among extreme ages below 20 (58.5%) and above30 (48.8%). Those having regular ANC follow up shows higher have less occurrence of LBW (70% vs. 20%) (Agarwal et al.  2011 ). Another cross-sectional study in Maternity hospitals in Iran 2008 overall prevalence as 8.8% with high prevalence (33.3%) among low maternal weight gain during pregnancy and low maternal BMI (Golestan M et al. 2011). In contrary Ghana study on 1200 samples showed 21.1% prevalence. A nationally representative, cross-sectional surveys in India shows the prevalence of LBW was 20.5% (95% CI: 19.5%, 21.4%) in 2005/2006 (P = 0.079). The prevalence of LBW was higher among younger mothers and those with a lower BMI (Yarlini et al. 2013).An institution based comparative cross-sectional study using consecutive sampling technique to assess the prevalence and associated factors of Low Birth Weight in Axum (urban) and Laelay Maichew (rural) districts, Tigray, north Ethiopia on 520 live birth neonates from both urban and rural district shows a low birth weight prevalence of 9.9% and 6.3% in Axum and Laelay Maichew districts, respectively (Teklehaimanot et al. 2014). A cross-sectional study in Jimma Ethiopia shows 22.5 % LBW prevalence. Mothers residing in the urban setting had higher risk of delivering LBW babies (p = 0.001) (Tema, 2006)**.** While other study in Jimma shows lower prevalence of LBW, 11.02% (Gebremariam, 2005)**.** While study conducted in Kersa Demographic Surveillance and Health Research Center (KDS-HRC) field site shows an incidence of LBW as 28.3% which may underestimate the actual prevalence (Nega et al., 2012)

## 2.2. International Literature on Low Birth Weight

Numerous studies have investigated LBW in various regions of the world. The results of those studies outlined here in order to illustrate the situation from both a developed and developing countries. Khatun, S., & Rahman, M. (2008) conducted a study in Bangladesh to analyze socioeconomic determinants of low birth weight using logistic regression analysis. A total of 1,467 births occurred during the study period, of which 465 met the study criteria. Among which one hundred and eight LBW babies were compared with 357 normal birth weight babies. Out of 20 possible risk variables analyzed, nine were found significant when studied separately. Mother's age, education, occupation, yearly income, gravid status, gestational age at first visit, number of antenatal care visit attended, quality of antenatal care received and pre-delivery body mass index had significantly associated with the incidence of LBW. Using the step wise logistic regression, mother's age ($p<0.001$), education ($p<0.02$), number of antenatal care visit attended ($p<0.001$, OR=29.386) and yearly income ($p<0.001$, OR=3.379) created the best model, which predicted 86.1% and 94.4% of the LBW babies and normal birth weight babies respectively. Maternal age, educational level and economic status play an important role in the incidence of low birth weight. Dharmalingam, *et al*., (2010) conducted a study from India using national survey data investigated the association between the mother's nutritional status and birth weight of her new born. The authors concluded that nutritional status, as measured by the mother's body mass index, was the most important determinant of LBW. Other important determinants included safe drinking water, antenatal care utilization, and anemia. Another study examined the association between social factors and newborn birth weight in a population in Quebec, Canada (Dubois, L., & Girard, M. 2006). Results demonstrated that birth weight increased with higher levels of family socioeconomic status and with higher maternal body mass indices. Newborn birth weight was lower among mothers who smoked. Body mass index was the most important indicator of LBW among mothers of higher socioeconomic status; however, maternal smoking was the most important indicator among mothers of lower socioeconomic status. Findings from these two studies may suggest that while many of the determinants of LBW may be similar in developed and developing countries, there are disparities reflective of local genetic, cultural, and environmental contexts. Brawarsky, P., *et al*., (2012) carried out a case-control study investigating risk factors for LBW in Sancti Spiritus, Cuba. Cases consisted of 764 singleton live births of less than 2,500grams while controls consisted of 1,437 singleton live births of at least

2,500 grams, selected from the same hospital. Data were obtained from clinical histories, birth registries, and personal interviews with the mothers. Multivariate analyses revealed an increased likelihood of LBW for mothers with anemia, with a gestational weight gain of less than eight kilograms, and for mothers who smoked during pregnancy. There was no association found between LBW and low educational attainment (incomplete primary school or less) or late attendance at first antenatal care visit.

**2.3 Literature Review of Low Birth weight in Africa**

Mwabu, G (2011) investigated the determinants of birth weight in Kenya in the year 2009 using data from welfare monitoring surveys collected by the Central Bureau of Statistics, Ministry of Planning and National Development. Structural equation model was used for analysis. It is shown that immunization of the mother against tetanus during pregnancy has a strong positive effect on birth weight. Other determinants of birth weight include age of the mother at first birth and birth orders of siblings. It is further shown that birth weight is positively associated with mother's age at first birth and with higher birth orders, with the first-born child being significantly lighter than subsequent children. Moreover, birth weights are lower in rural than in urban areas and females are born lighter than males. There is tentative evidence that babies born at the clinics are heavier than babies from the general population. Siza J.E. (2008) carried out a descriptive retrospective cross - sectional study investigating the risk factor associated with LBW using existing data from a one-year (2006) block of birth registers of 3464 pregnant women was done at Kilimanjaro Christian Medical Centre in Moshi, Tanzania. HIV positive women were twice more likely to give birth to LBW infants than HIV negative ones ($\chi2 = 6.7$; $P<0.01$; OR = 2.4; 1.1, 5.1). Mothers without formal education were 4 times more likely to give birth to LBW neonates than those who had attained higher education (OR= 3.6; 2.2, 5.9). There was a linear decrease in low birth weights of newborns as maternal educational level increased ($\chi2$ for linear trend = 42.7; $P< 0.01$).There was no statistically significant difference among parent's occupations regarding LBW of their newborns. Unmarried mothers were more likely to give birth to LBW neonates as compared to their married counterparts (OR = 1.65; 1.2, 2.2) and the difference was statistically significant ($\chi2=13.0$, $P< 0.01$). Hypertension, pre-eclampsia and eclampsia disease complex had the highest prevalence (46.67%) and population attributable fraction of low birth weight (PAF = 25.2%; CI= 22.0-27.6). Bleeding and schistosomiasis had the same prevalence (33.33%) of LBW babies. Other complications and diseases that contributed to high prevalence of LBW included anemia (25%), thromboembolic diseases (20%), tuberculosis (17%) and malaria (14.8%). LBW was strongly associated with gestational age below 37 weeks (OR = 2.00; CI=1.5, 2.8) contributing to 42% of LBW deliveries in the study population (PAF = 42.4%: 25, 55). Pregnant women with malnutrition (BMI<18) gave the highest proportions 17% of LBW children followed by underweight (BMI; 18-22) who gave15.5% of LBW neonates. There was statistically significant difference between the proportions of LBW infants from mothers who did not receive antenatal care (28.6%) and those who attended for the services (13.8%) ($\chi2 = 8.8$; $P= 0.01$).Ipadeola, O. B., *et al*., (2013) examine the influence of household poverty levels and maternal nutritional status on child's weight at birth using 2008 Nigeria Demographic Health Survey (NDHS) measures weight at birth on an ordinal scale. Ordinal logistic regression technique was employed for all analyses. Quintiles of wealth index were used as a measure of assets owned by households while body mass index was used to assess maternal nutritional status. Other demographic characteristics such as mother's age at birth of the child, educational attainment, locality (urban/rural) and geo-political zones were controlled for in the models. The sample size for survey was 5138. Wealth index and maternal nutritional status were positively associated with child's weight at birth, while mother's educational attainment was not statistically significant. Significant and positive association of wealth index was evident with middle (OR=1.38, $p<0.0001$), higher (OR=1.48, p= $p<0.0001$), and highest (OR=1.37,p=0.009) when compared with those in the poorest category of wealth index. Mothers that were too thin or underweight based on their BMI, were more likely to give birth to children with low birth weight. (OR=0.80, p=0.008); while those that weighed more than they should (overweight: OR=1.35, $p<0.0001$; or obese: OR=1.29, p=0.065) were more likely to give birth to children with large weights when compared with mothers with normal BMI. Significant gender differentials were also found. Males were about 1.4 times ($p<0.0001$) more likely to have weights larger than their female counterparts at birth. Age of mother at the birth of a child has also been shown to be of risk to pregnancy outcomes. Teenage mothers were more likely to give birth to children with low birth weight. Here, positive significant association was observed for mothers' age at birth and child's weight at birth. Children from mothers in the age range 25 to 39 years were about 1.26 times more likely to weigh more at birth compared with children from teenage mothers ($p<0.05$). Significant spatial pattern was observed at the level of geopolitical zones with $p<0.05$. This spatial variation, however, needs to be investigated further at a highly disaggregated level of states as information at this level could be masked. Multiple births are significantly associated with low birth weight compared with singleton births (OR=0.59, p<0.0001).

### 2.4 Literature Review of Low Birth Weight in Ethiopia

Tema (2006), Conducted a cross-sectional descriptive study to assess the Prevalence and determinants of low birth weight in Jimma Zone, Southwest Ethiopia. Mothers with new born delivered in the four health centers (Jimma, Agaro, Asendabo and Shebe) and Jimma university hospital from September 1, 2002 to march 30, 2003, and those delivered at home and received care within the first 24 hours after delivery in these health care settings. A total of 145 (22.5%) of the newborns were LBW. Mothers residing in the urban setting had higher risk of delivering LBW babies and the difference was statistically significant (p = 0.000).Analysis of maternal obstetric history revealed that those mothers who delivered before 37 weeks of gestation, had weight loss, and who did not receive additional diet during pregnancy had higher risk of delivering LBW babies and the difference was statistically significant (p =0.01, 0.00, 0.00) respectively. Similarly, those who had multiple gestation had a higher risk of delivering LBW babies and the difference was statistically significant (p = 0.00).In general, therefore, the literature investigating LBW from the above studies have found several determinants that increase the likelihood of delivering a LBW infant. These include smoking during pregnancy, low gestational weight gain, inadequate antenatal care, low educational attainment, less skilled occupation, maternal pre-pregnancy weight, low gestational weight gain, anemia and female sex of the newborn. Few studies have found that higher calorie and protein reserves (i.e. the mother’s nutritional status) had a positive effect on infant birth weight, concluding that the mother’s nutritional status is a key determinant of newborn birth weight (Karim E, Mascie-Taylor 2012)

## 2.5 Literature Review of Multilevel Regression Model

Traditional logistic regression tended to increase the statistical significance for the effects of variables measured at the higher-level compared to the level of significance indicated by the multilevel model. (Austi and Alte, 2003).The multilevel regression model has become known in the research literature under a variety of names, such as 'random coefficient models, 'variance component models, and 'hierarchical linear model. Statistically oriented publications tend to refer to the model as a mixed-effects or mixed model. The multilevel regression models assume that there is a hierarchical data set, with one single outcome or response variable that is measured at the lowest level, and explanatory variables at all existing levels. Conceptually, it is useful to view the multilevel regression model as a hierarchical system of regression equations (Joop, 2010).

Multilevel analysis originally developed in the fields of education, sociology, and demography, has received increasing attention in other fields like public health and epidemiology over the past few years. Mason et al (1983) were among the first to develop the concepts and methodology for analyzing multilevel data. Further methodological and substantive work by Bryk et al (1992) and Goldstein (1987, 1995, and 2003) has popularized the multilevel models for linear data. Multilevel models and their generalizations to categorical outcomes in areas of research including biostatistics and economics have been important and active areas of statistical research. Stiratelli et al (1984) and Mason et al (1985) were among the first to deal with binary outcomes. Earlier other methodological work on multilevel logistic models includes Anderson (1985), Conaway (1989), Goldstein (1991), Bryk et al (1993), Long ford (1993) and Ng et al (2006). Several review articles, for example, Rodriguez et al (1995) and Prendergast et al (1996) have discussed and compared some of these models and their estimation procedures. Also, Snijders et al (1999), J. Hox (2002) and Gelman et al (2007) provide a practical summary of the multilevel logistic regression model and the various procedures for estimating its parameters.

Many statisticians and social scientists have been using multilevel logistic regression models for analyzing binary data since its development. Finally we have analyzed the 2011 EDHS low birth weight binary data using hierarchical logistic modeling technique. The treatment is at the groups (region) level, but the outcome is measured on individual infants. Here the units at lower level (level-1) are individuals (infants) who are nested within units at higher level (level-2) are households (families) and the households are again nested within units at the next higher level (level-3) are regions. The fact that the regional states in Ethiopia had a variety of socio-economic and demographic factors to encourage low birth weight of infants at their region and national level. Indeed, not only regional-level differentials but also there are the individual-level factors attributed for low birth weight infants in addition to demographic factors of children as well. This differential among individual, region, national and also through continent level indicated the facts that, the rate of low birth weight in developed and developing country has different structure. The response variable in this study is “low birth weight infants” which is binary and hence multilevel logistic regression model is a natural choice for modeling. The multilevel logistic regression analysis considers the variations due to hierarchy structure in the data. It allows the simultaneous examination of the effects of group level (household and region) and individual level variables on individual level outcomes while accounting for the non-independence of observations within groups.

# Chapter Three

## 3. Data and Methodology

### 3.1 Data

The source of data for this study is the Ethiopian

Demographic and Health Survey (EDHS), which was obtained from Central Statistical Agency (CSA).Its objective is to collect information and estimate of some of the MDG indicators such as childhood mortality, knowledge and use of family planning methods, maternal and child health, nutrition, knowledge of HIV/AIDS were provided for the nine regional states and two city administrations.

### 3.2 Description of study area

The 2007 Population and Housing Census, conducted by the CSA, provided the sampling frame from which the 2011 EDHS sample was drawn. Administratively, regions in Ethiopia are divided into zones, and zones into administrative units called woreda's. Each woreda is further subdivided into the lowest administrative unit, called kebele. During the 2007 Census, each Kebele is subdivided into census enumeration areas (EAs) or clusters, which are convenient for the implementation of the census.

### 3.3 Sampling Design

The 2011 EDHS sample was selected using a stratified, two-stage cluster sampling design. Clusters were the sampling units for the first stage. The sample included 624 clusters, 187 in urban areas and 437 in rural areas. Households comprised the second stage of sampling. In the second stage, a fixed number of 30 households were selected for each cluster. A complete listing of households was carried out in each of the selected clusters from September 2010 through January 2011 (CSA, 2011).

### 3.4 Variables

#### 3.4.1 Response variable

The response variable of the study is the low birth weight of infants. The outcome of interest is a binary variable such as small birth weight versus large birth weight. The child weight is first dichotomized based on the cut-off points as described in literature review leading to the binary response (UNICEF/WHO, 2004).

Table 3.1: Coding and explanation of response variable

| Variable | Presentation of variable | Factor coding |
|---|---|---|
| Child weight at Birth | Child weight | 1=low birth weight(<2500 g)<br>0=High birth weight(≥2500 g) |

### 3.4.2 Explanatory variables

Table 3.2 the predictors assessed as the main determinants of low birth weight in this study are described as follows. Descriptions and coding of the study variables.

| Attribute | Description | Categories |
|---|---|---|
| 1.Region | Regions of Ethiopia | 1. Tigray<br>2. Afar<br>3. Amhara<br>4. oromia<br>5. somali<br>6. Beneshangul-Gumuz<br>7. SNNP<br>8. Gambela<br>9. Harrari<br>10. AddisAbaba<br>11.Dire Dawa |
| 2. Sex | Sex of child | 0= male 1=female |
| 3. Residence | Maternal place of residence | 0=urban 1=Rural |
| 4. Wealth status | Maternal income level | 0=poorest 1= poorer 2=middle 3. richer 4.richest |
| 5.age | Maternal age | 0=10-19 1=20-39 2=40-49 |
| 6.antenatal visits | Number of antenatal visit during pregnancy | 0=visit<br>1=No visit |
| 7.Birth | Multiple birth | 0=single 1= multiple |
| 8.Anemia | Maternal anemia | 0=sever 1=moderate 2=mild<br>3=not anemic |
| 9. Education level | Maternal education level | 0=No education 1=primary<br>2=secondary 3=higher |
| 10. Height | Maternal height | continuous |
| 11. Weight | Maternal weight | continuous |

## 3.5 Methodology

### 3.5.1 Generalized linear model (GLM)

Generalized linear models (GLMs) extend ordinary regression models to encompass non normal response distributions and modeling functions of the mean (Agresti, 2002). Three components that specify a generalized linear model are random component, which identifies the response variable Y and its probability distribution; a systematic component specifies explanatory variables used in a linear predictor function; and a link function specifies the function of expected value of the response variable that the model equates to the systematic component.

### 3.5.2 Logistic Regression for Binary Data

Many of the analyses in this thesis utilize logistic regression due to there being a binary response variable, indicating the presence or absence of a specific factor.Logistic regression is a popular modeling approach when the dependent variable is dichotomous. This model allows one to predict such outcomes, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Hosmer et al (2000) has described logistic regression focusing on its theoretical and applied aspect.

A binary outcome for the $i^{th}$ individual is denoted as $y_i$ , where $y_i = 0 \; or \; 1$.The probability that $y_i = 1$ is given as $p_i$. Let E (Y | x) = $p_i$ be the conditional mean of dependent variable, Y given explanatory variable, x. Then for *K* explanatory variables, denoted for each individual as $x_{ki}$ (where *k* = 1, ..., K *)*, the general model for a binary response is

$$f(p_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} \qquad (3.1)$$

Where $f(p_i)$ is some transformation of $p_i$. This transformation is required as the range for $p_i$ is (0, 1), as it represents a probability, and thus the simple application of a linear model may produce probabilities where $p_i > 0$ or < 1. A function is chosen, called the link function, which transforms the $p_i$ to have a range $(-\infty, \infty)$. There are a number of choices for the link function, but the most widely used due to ease of interpretation is the logit transformation. The logit transformation is shown below:

$$f(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$$

Where $(\frac{p_i}{1-p_i})$ is the odd of $y_i = 1$.Using the logit link function, the model for binary data is given as:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} \qquad (3.2)$$

To obtain the odds that $p_i$, exponentials of each side of the equation in (3.2) are taken. To obtain $p_i$ , the expression that is required is:

$$p_i = \frac{\exp(\beta_0+\beta_1 x_{1i}+\cdots+\beta_k x_{ki})}{1+\exp(\beta_0+\beta_1 x_{1i}+\cdots+\beta_k x_{ki})} \qquad (3.3)$$

The logistic regression model is easily solved by the Eq. 3.3.The quantity $\frac{p_i}{1-p_i}$ is called odds and hence the logit is called log odds. There are two odds: one is when Y = 1 and the other is when Y = 0. The ratio of these two odds is known as odds ratio denoted by $\psi$ which is the base for interpretation of the coefficients of the logistic regression model. The $\psi$ is the probability that Y will be a member of one class relative to the other class. For instance, for a binary independent variable (x: 0 or 1), $\psi$ can be expressed as below,

$$\Psi = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)} = e^{\beta_1}$$

This odds ratio is the measure of how much more likely or unlikely it is for the outcome to be present among those with x = 1 than among those with x = 0. Hence, after estimating the parameters the effect of the independent variable on outcome variable can be measured through this odds ratio. The exponential of each coefficient $\beta_k$, is interpreted as an odds ratio which will give the effect of a one-unit increase in $x_k$ on the odds that $y_i = 1$, ceteris paribus.

### 3.5.3 Multilevel Linear Model

The multilevel linear model and its application had been described by various authors in the past [Mason et al (1983), Goldstein (1987, 1995, 2003 ), Bryk et al (1992)]. We describe below the multilevel linear model and its basic properties. We first consider a simple linear model for the data with hierarchical structure (with two levels) with a single explanatory variable,

$$y_{ij} = \alpha_{0j} + \beta_1 x_{ij} + e_{ij} \tag{3.4}$$

Where $y_{ij}$ is the outcome variable for the $i^{th}$ unit at level-1 and the $j^{th}$ unit at level-2, $\alpha_{0j}$ is the intercept for the $j^{th}$ unit at level-2 (i.e. it varies across level-2 ), $x_{ij}$ is the explanatory variable for the $i^{th}$ unit at level-1 and the $j^{th}$ unit at level-2, $\beta_1$ is the effect of $x_{ij}$ and $e_{ij}$ is the level-1 random effect. Here, $\alpha_{0j}$ is a random variable rather than a constant and can be written as:

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

(3.5)

Where, $\beta_0$ is the intercept (constant across level-2) and $u_{0j}$ is a random effect accounting for the random variation at level-2. Combining both equations (3.4) and (3.5) the two level linear model can be written as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij} \tag{3.6}$$

In equation (3.6), $u_{0j}$ and $e_{ij}$ are random quantities which follow normal distributions, $N(0, \sigma_{u0}^2)$ and $N(0, \sigma_e^2)$ respectively. The equation (3.6) has the properties: $E[u_{0j}] = E[e_{ij}] = 0\,, v(u_{0j}) = \sigma_{u0}^2, \quad (e_{ij}) = \sigma_e^2 \quad, cov(u_{0j}, e_{ij}) = 0\ and\ cov(u_{0j}, u'_{0j}) = 0\ for\ j \neq j'$. In this model, $\beta_0$ and $\beta_1$ are known as fixed parameters and $\sigma_{u0}^2$ and $\sigma_e^2$ are known as random parameters. Equation (3.6) is also known as variance component model since the variance of the response, about the fixed components $(\beta_0 \& \beta_1)$, is

$$\mathrm{var}(y_{ij}/\beta_0, \beta_1, x_{ij}) = \mathrm{var}(\mathrm{u_{0j}} + \mathrm{e_{ij}}) = \sigma_{\mathrm{u0}}^2 + \sigma_{\mathrm{e}}^2$$

Which is the total variation obtained summing level 1 and level 2 variance. The covariance between two units of level 1 $(say, i_1, i_2)$ can be defined as,

$$cov(u_{0j} + e_{i1j}, u_{0j} + e_{i2j}) = \sigma_{u0}^2$$

The within level-2 or intra-level 2 correlations after controlling the explanatory variable can be obtained from:

$$\rho = \frac{\sigma_{u0}^2}{(\sigma_{u0}^2+\sigma_e^2)}$$

This two level model (3.6) can be extended to a three level model with random coefficient by the following equation:

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + v_{1k}x_{ijk} + u_{1jk}x_{ijk} + v_{0k} + u_{0jk} + e_{oijk}$$

where k indexes level 3, $v_{0k}$ and $u_{0jk}$ are the random intercepts for level 3 and level 2 respectively, $x_{ijk}$ is an observed explanatory variable , $u_{1jk}$ is $x_{ijk}^{'s}$ random effect at level 2 and $v_{1k}$ is $x_{ijk}^{\prime s}$ random effect at level 3. Other parameters of the above model include $E[v_{0k}] = E\left[u_{0jk}\right] = E\left[e_{0ijk}\right] = 0\ , var(v_{0k}) = \sigma_{v0}^2 . var\left(u_{ojk}\right) = \sigma_{uo}^2, var\left(u_{1jk}\right) = \sigma_{u1}^2, var(v_{1k}) = \sigma_{v1}^2, var\left(e_{0ijk}\right) = \sigma_{e0}^2\ and\ cov\left(u_{0jk}, u_{1jk}\right) = \sigma_{uo1}$ The total variance and its partition for this three level model can be easily found [for details see Goldstein (2003)].

## 3.6 Multilevel Logistic Regression Model

We shall start considering first a two level logistic regression model with a single explanatory variable. Then a three level model with both fixed and random effect.

**Two Level Model**

Basically, the two level logistic models is equivalent to model (3.6) except for the outcome variable. Let $y_{ij}$ be the binary outcome variable, coded ‘0’ or ‘1’, associated with level-1 unit i nested within level-2 unit j. Also let $p_{ij}$ be the probability that the response variable equals 1, and $p_{ij} = pr(y_{ij} = 1)$. Here, $y_{ij}$ follows a Bernoulli distribution. Like logistic regression the $p_{ij}$ is modeled using link function, logit. The two level logistic regression model can be written as,

$$\ln\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_0 + \beta_1 x_{ij} + u_{0j} \tag{3.7}$$

Where $u_{0j}$ is the random effect at level 2. Without $u_{0j}$, Eq. (3.6) can be considered as a standard logistic regression model. Therefore, conditional on $u_{0j}$ , the $y_{ij}^{'s}$ can be assumed to be independently distributed. Here, $u_{0j}$ is a random quantity and follows $N(0, \sigma_{u0}^2)$. The model

(3.6) can be written as follows splitting up into two models: one for level 1 and the other for level 2.

$$\ln\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{0j} + \beta_1 x_{ij} \qquad \text{[Model: level 1]}$$

$and$

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

The multilevel logistic regression model cannot be derived in the way simple logistic regression model is derived. This model (3.6) can be derived through a latent or hidden variable conceptualization. Let us suppose $y'_{ij}$ to be a continuous variable such that

$$y_{ij} = 0 \qquad if\ y'_{ij} \le 0$$

$and$

$$y_{ij} = 1 \qquad if y'_{ij} > 0$$

However we cannot observe $y'_{ij}$ directly but only the binary outcome $y_{ij}$ . In terms of the continuous latent variable $y'_{ij}$ , the model can be written equivalently to (3.6) as below,

$$y'_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij} \qquad (3.8)$$

Conditional on the random effect $u_{0j}$ at level two, a multilevel logistic model can be derived from (3.7) depending on the standard logistic distribution of $e_{ij}$ . This conceptualization or threshold concept illustrates the close connections between the multilevel models for linear data and those for binary data [McCullagh &Nelder (1989)]. Conditional on $u_{0j}$ , the conditional density function for cluster j for model (3.6) is identical to that for the logistic regression

$$f\left(y_j|x_j, u_{0j}\right) = \prod_{i=1}^{n_j} \frac{\exp\left[y_{ij}(\beta_0+\beta_1 x_{ij}+u_{0j})\right]}{1+\exp\left(\beta_0+\beta_1 x_{ij}+u_{0j}\right)} \qquad (3.9)$$

Where, $y_j$ and $x_j$ denote the responses and explanatory variables in cluster j respectively.

### Three Level Models

For three levels the logistic regression model with no explanatory that has both a fixed effect and a random effect can be written as,

$$\ln\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = \beta_0 + v_{0k} + u_{0jk} \quad (3.10)$$

*or*

$$logit\left[pr(y_{ijk} = 1|v_{0k}, u_{ojk})\right] = \beta_0 + v_{0k} + u_{0jk} \quad (3.11)$$

Where i, j and k denote, respectively, levels 1, 2 and 3; $v_{ok}$ and $u_{0jk}$ are the random intercepts for level 3 and level 2 respectively. Here, $v_{ok} \sim N(0, \sigma^2)$ and $u_{ojk} \sim N(0, \tau^2)$, where $\sigma^2$ is the variance of the random intercept for level 3 and $\tau^2$ is the variance of the random intercept for level 2.

For three levels the logistic regression model with a single explanatory variable that has both a fixed effect and a random effect can be written as,

$$\ln\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = \beta_0 + \beta_1 x_{ijk} + v_{1k}x_{ijk} + u_{1jk}x_{ijk} + v_{0k} + u_{ojk}$$
(3.12)

Or

$$logit\left[\Pr(y_{ijk} = 1|v_{0k}, u_{0jk}, u_{1jk}, v_{1k})\right] = \beta_0 + \beta_1 x_{ijk} + v_{1k}x_{ijk} + u_{1jk}x_{ijk} + v_{ok} + u_{ojk}$$
(3.13)

Model (3.12) can be written, by splitting up, for each level as below,

$$\ln\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = \beta_{ojk} + \beta_{1j}x_{ij} \qquad \text{[Model: level 1]}$$

$$\beta_{0jk} = \beta_{0k} + u_{0jk} \qquad \text{[Model: level 2]}$$

$$\beta_{1j} = \beta_{1k} + u_{1jk} \qquad \text{[Model: level 2]}$$

$$\beta_{0k} = \beta_0 + v_{ok}$$

$$\beta_{1k} = \beta_1 + v_{1k} \qquad \text{[Model: level 3]}$$

Similarly, this model can be extended for level more than three incorporating the fixed or random or both components in the model.

### 3.6.1 The Empty Model

We first estimated a model with no predictors i.e. an intercept-only model (Empty model) that predicts the probability of low birth weight of infants. The multilevel model is described by the following equations.

$$Logit(p_{ijk}) = \ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \beta_{0jk} \qquad \text{(Model: level 1)}$$
$$\beta_{0jk} = \beta_{0k} + u_{0jk} \qquad \text{(Model: level 2)}$$
$$\beta_{0k} = \beta_0 + v_{ok} \qquad \text{(Model: level 3)}$$

Where, $v_{0k} \sim N(0, \sigma^2)\ and\ u_{ojk} \sim N(0, \tau^2)$

This model does not include a separate parameter for the individual level variance. This is because the individual level residual variance of the $Y_{ijk}$ (low or high birth weight infants) follows Bernoulli distribution directly from the probability of having low birth weight infants $p_{ijk}$.

### 3.6.2 Random Intercept multilevel Logistic Regression Mode l

The earlier intercept model analysis investigates how standard logistic model differs from a multilevel model when no explanatory variable is considered in the model. Also it checks how the estimates vary across different estimation methods. Now in this random intercept multilevel analysis we will find out whether each of the explanatory variables is influencing the response 'LBW' and how much the estimates distorted from the actual when multilevel effect has not considered. In the multilevel analysis each of the models presents a random intercept and a fixed slope for the variable.

Consider the explanatory variable $x_{ijk}$, the probability of having low birth weight depend on indicators was denoted by $p_{ijk}$. Then, random intercept model expresses the log odds, i.e. the logit of $p_{ijk}$ , is the sum of a linear function of all indicators of low birth weight is given as

$$Logit(p_{ijk}) = \ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \beta_{0jk} + \beta_1 x_{1ijk} \qquad (3.15)$$

Where, $Logit(p_{ijk})$ does not include a level-one residual because it is an equation for the probability of having low birth weight infants $p_{ijk}$ rather than for the out come $y_{ijk}$.

$\beta_{0jk}$- is assumed to vary randomly and $\beta_0$ is given by the sum of an average intercept and $v_{0k}$, $u_{0jk}$ are group (region) dependent deviations and household dependent deviations respectively is given by:

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$

By replacing it in equation (3.14), we have

$$Logit(p_{ijk}) = \beta_0 + \beta_1 x_{ijk} + v_{0k} + u_{0jk}$$

Or

$$p_{ijk} = \frac{\exp\,(\beta_0+\beta_1 x_{ijk}+v_{0k}+u_{0jk})}{1+\exp\,(\beta_0+\beta_1 x_{ijk}+v_{0k}+u_{ojk})} \qquad (3.16)$$

Where, $\beta_1$- is a unit difference between the values of two individuals in the same group is associated with a difference of in $\beta_1$ their log odds, or equivalently, a ratio of exp $(\beta_1)$ in their odds.

$u_{0jk}$ -is random part of the model and It is assumed that they are mutually independent and normally distributed with mean zero and variance $\delta^2_{0u}$.

$v_{0k}$ -is random part of the model and it is assumed that they are mutually independent and normally distributed with mean zero and variance $\delta^2_{0v}$.

### 3.6.3 Random slope multilevel logistic regression model

Earlier we considered fitting intercept-only model with random intercept and then multilevel models with random intercept for each predictor separately. Now this univariate multilevel model with random intercept can be expanded into an univariate multilevel model with random slope. The variance components models that we have just worked with assume that the only variation between households or regions is in their intercepts i.e. the slope was fixed for all households and for all regions. But there is the possibility that the regions have different slopes and also the households have different slopes i.e. the probability of low birth weight varies across both households and regions. This implies that the coefficient of each explanatory variable will vary from region to region and from household to household.

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

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$
$$\beta_{1jk} = \beta_1 + v_{1k} + u_{1jk}$$

Thus, by substituting in equation (3.15) then, $logit(p_{ijk})$ is given as:

$$\begin{aligned} Logit(p_{ijk}) &= (\beta_0 + v_{0k} + u_{0jk}) + (\beta_1 + v_{1k} + u_{1jk})x_{ijk} \\ &= \beta_0 + \beta_1 x_{ijk} + v_{0k} + u_{0jk} + v_{1k}x_{ijk} + u_{1jk}x_{ijk} \end{aligned} \qquad (3.17)$$

Now, we have two random effects at region level and household level, the random intercept $v_{0k}$ and the random slope $v_{1k}$ and $u_{0jk,}$random intercept, random slope $u_{1jk}$ respectively. It assumed that both random effects have mean zero. And the variances are denoted by $\delta_{0v}^2$ , $\delta_{1v}^2, \delta_{0u}^2, \delta_{1u}^2$ , $\delta_{01v}^2$ and $\delta_{01u}^2$ their covariance. Where, $\beta_0$ -is the average intercept of the response variable. $\beta_1$ -is fixed regression coefficient given explanatory variable.

Now, we are going to extend the above single explanatory model by including more explanatory variable that has random effects on outcome variables. Suppose that there are H level-one explanatory variables$X_1, X_2, \ldots, X_H$, and consider the model where all predictor variables have varying slopes and random intercept

That is:

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}\ , \beta_{1jk} = \beta_1 + v_{1k} + u_{1jk}, \dots, \beta_{hjk} = \beta_h + v_{Hk} + u_{hjk}$$ , for

h=1, 2,…,H, then we have:

$$Logit(p_{ijk}) = (\beta_0 + v_{0k} + u_{0jk}) + (\beta_1 + v_{1k} + u_{1jk})x_{1ijk} + \dots + (\beta_h + v_{Hk} + u_{hjk})x_{hijk}$$

$$= \beta_0 + \sum_{h=1}^{k} \beta_h x_{hijk} + v_{0k} + u_{0jk} + \sum_{1}^{H} v_{Hk} x_{Hijk} + \sum_{h=1}^{H} u_{hjk} x_{hijk} \qquad (3.18)$$

Where, $\beta_0 + \sum_{h=1}^{k} \beta_h x_{hijk}$ -is fixed part of the model and $v_{0k} + \sum_{h=1}^{H} v_{hk} x_{hijk} + u_{0jk} + \sum_{h=1}^{k} u_{hj} x_{hijk}$ -is the random part of the model.

## 3.7 Parameter Estimation Methods

The most common methods for estimating multilevel logistic models, used in this study, are based on likelihood. Among the methods, Marginal Quasi Likelihood (MQL) (Goldstein, 1991; Goldstein and Rasbash, 1996) and Penalized Quasi Likelihood (PQL) (Laird, 1978; Breslow and Clayton, 1993) are the two prevailing approximation procedures. Rodriguez et al (1997) compared four approximation estimation procedures (first-order $MQL^1$, second-order $MQL^2$, first-order PQL, and second-order PQL) with the likelihood achieved through high-dimensional numerical integration and the method of Gibbs sampling. They found that the second-order MQL and PQL were producing more accurate estimates than the first-order ones because they used some of the second-order terms in the Taylor expansion. Finally they concluded that all approximation methods (MQL-1, MQL-2, PQL-1, and PQL-2) underestimate the random as well as fixed effects and that the underestimations of MQL-1, MQL-2, and PQL-1 are severe. They preferred PQL-2 to all other methods as it has been found least biased. Since their research in 1997 it has been a norm to prefer the PQL-2 method as a multilevel estimation technique for binary data in many socio-economic and demographic studies when the estimates across other methods vary significantly [Goldstein (2003 In the context of our analysis throughout the study we also preferred PQL-2 method to all other methods including MCMC because PQL-2 model produces estimates closer to the true values [also see Goldstein (2003)]. After applying these quasi likelihood methods, the model is then estimated using iterative generalized least squares (IGLS) or reweighted IGLS (RIGLS) (Goldstein, 2003). Second-order PQL method has been used throughout the multi-level analyses since this method approximates well compared to the other PQL and MQL methods (Goldstein, 2003). Details of the PQL method are given below.
Bayesian methods using Markov chain Monte Carlo (MCMC) have also been used for parameter estimation.

**Penalized quasi-likelihood**

The PQL estimation procedure is described here for two level logistic regression models. Consider a level-1 outcome $y_{ij}$ taking on a value of 1 with conditional probability$p_{ij}$. Then the logit model or the generalized linear model is,

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \eta_{ij} = X_{ij}^T + Z_{ij}^T U_{oj}$$

for level-1 unit *i* nested within level-2 unit $j$. At level 1, we assume $y_{ij}$ conditionally distributed as Bernoulli, while the random effects vector $U_{oj}$ is distributed as $N(0, \sigma_u^2)$ across the level-2 units. Let us consider the variance $\sigma_u^2$ as $T$ throughout this PQL estimation procedure. The PQL approach can be derived as a nonlinear regression model. In the case of binary outcomes with logit link, we start with the level-1 model

$$y_{ij} = p_{ij} + e_{ij} \tag{3.19}$$

Where$E(e_{ij}) = 0$ *and* $Var(e_{ij}) = p_{ij}(1 - p_{ij})$. This is a nonlinear model which we linearize by means of the first-order Taylor series expansion. At these iterations, we have

$$p_{ij} \backsimeq p_{ij}^s + \frac{dp_{ij}}{d\eta_{ij}}(\eta_{ij} - \eta_{ij}^s)$$

and evaluate the derivative

$$\frac{dp_{ij}}{d\eta_{ij}} = p_{ij}(1 - p_{ij}) = w_{ij}\text{ , at } p_{ij}^s$$

Substituting the linear approximation for $p_{ij}$ in equation (3.18) yields

$$y_{ij} = p_{ij}^s + w_{ij}^s(\eta_{ij} - \eta_{ij}^s) + e_{ij}$$

Algebraically rearranging this equation so that all known quantities are on the left-hand side of the equation produces

$$\frac{y_{ij}-p_{ij}^s}{w_{ij}^s} + \eta_{ij}^s = \eta_{ij} + \frac{e_{ij}}{w_{ij}^s}$$

This equation has the form of the familiar two-level hierarchical linear model

$$y_{ij}^{*(s)} = X_{ij}^T\gamma + Z_{ij}^T U_{oj} + \varepsilon_{ij}$$

This gives a straightforward updating scheme. This is known as penalized quasi-likelihood because it is obtained by optimizing a quasi-likelihood (involving only 1st and 2nd derivatives) with a penalty term on the random effects. Here,

$$y_{ij}^{*(s)} = \frac{(y_{ij} - p_{ij}^{s})}{w_{ij}^{s} + \eta_{ij}^{s}}, \qquad \varepsilon_{ij} = \frac{e_{ij}}{w_{ij}^{s}} \sim N\left(0, w_{ij}^{(s)-1}\right) and\ U_{oj} \sim N(0, T)$$

The estimate of $\eta_{ij}^{s}$ can be written as below

$$\eta_{ij}^{s} = X_{ij}^{T}\hat{\gamma}^{s} + Z_{ij}^{T}U_{oj}^{*(s)},$$

Where $U_{oj}^{*(s)}$ is the approximate posterior mode, i.e.

$$U_{0j}^{*(s)} = \left(Z_j^T w_j^{(s)} Z_j + T^{s-1}\right)^{-1} Z_j^T w_j^{s} (y_j^{*(s)} - X_j \hat{\gamma}^{s}) \ \ for\ \ w_j^{s} = dia\{w_{ij}^{s}, \dots, w_{nij}^{s}\}$$

## Results and Discussions

### 4.1 Results

The independent variables shown in chapter three (Table 3.2) supports the contention that they are influential and they have strong association with use of low birth weight. The primary choice of independent variables for this study was based on previous other studies on the factors influencing low birth weight [Dickute, J. and padaiga, z, et al,(2012), Tuntiseranee, p, et al(2013), Hirve, ss.Ganatra BR,(2008)].

Table 4.1 percentage of low birth weight (size of child) by predictors (covariates) results are based on our study sample=3715)

| Covariates | Measurement range | Birth weight | |
|---|---|---|---|
| | | Low | High |
| Sex | 0-1 | | |
| Male | | 47.1 | 52.9 |
| Female | | 59.6 | 40.4 |
| Wealth Index | 0-4 | | |
| Poorest | | 62.4 | 37.6 |
| Poorer | | 53.9 | 46.1 |
| Middle | | 53.1 | 46.9 |
| Richer | | 43.3 | 56.7 |
| Richest | | 43.4 | 56.6 |
| Place of residence | 0-1 | | |
| Urban | | 46.8 | 53.2 |
| Rural | | 54.3 | 45.7 |
| Education level | 0-3 | | |
| No education | | 56.5 | 43.5 |
| Primary | | 44.3 | 55.7 |
| Secondary | | 39.7 | 60.3 |
| Higher | | 35.3 | 64.7 |
| Anemia level | 0-3 | | |
| Sever | | 67.2 | 32.8 |
| Moderate | | 60.0 | 40.0 |
| Mild | | 56.4 | 43.6 |
| Not anemic | | 51.4 | 48.6 |

From table 4.1 results the total of 3715 children (0-59 months old) from nine regional states and two city administrations in Ethiopia were eligible for this study. Among these eligible regions, 1737 (46.8%) children were born with large weight whereas 1978 (53.2%) were born with small weight. The proportion of LBW is slightly larger (59.6%) for female child than the male child (47.1%). LBW is higher (62.4%) for poorest mothers when compared to mothers with poorer (53.9%), middle wealth status (53.1%), richer (43.3) richest mothers (43.1%).There is also a variation of LBW due to place of residence of mothers. The proportion of bearing child with LBW for rural mothers is (54.3%) and who living in urban area is (46.8%).Educational level of mothers has decreasing proportion to LBW. The proportion of LBW is (56.5%) for non-educated mothers, (44.3%) for primary educated mothers and (39.7%) for mothers whose education level is secondary (39.7%) and higher education (35.3%). Mothers who are not anemic (51.4%) have less proportion of bearing child with LBW than mothers who are mild anemic (56.4%), moderately anemic (60.0%) and severely anemic (67.2%).

From the 4.2 all explanatory variables are found highly statistically associated with dependent variable except maternal marital status, maternal preceding birth interval, maternal age, maternal weight and maternal height.

Table 4.2: Test of association: chi-square tests of independence between explanatory and dependent variable

| Independent variable | $\boldsymbol{x^2}$-significance | Independent variable | $\boldsymbol{x^2}$-significance |
|---|---|---|---|
| Multiple birth | $x_2^2$=12.959 (p<0.002) | Maternal Education level | $x_3^2$=48.411 (p<0.000) |
| Sex of child | $x_1^2$=58.657 (p<0.000) | Maternal anemia level | $x_3^2$=17.282 (p<0.0.01) |
| Maternal income | $x_4^2$=86.915 (p<0.000) | Preceding birth interval | $x_{139}^2$=142.06 (p<0.412) |
| Maternal age | $x_{33}^2$=27.795 (p<0.724) | Antenatal visit | $x_{17}^2$=91.459 (p<0.000) |
| Maternal residence | $x_1^2$=10.01 (p<0.002) | Marital status | $x_5^2$=6.667 (p<0.247) |
| Maternal weight | $x_{446}^2$=441.244 (p<0.295) | Maternal height | $x_{360}^2$=350.994 (p<0.623) |
| Region | $x_{10}^2$=207.617 (p<0.000) | Household | $x_{868}^2$=874.386 (p<0.433) |

We will also modeling how and to what extent all these explanatory variables are statistically related with low birth weight later in this chapter. MLwiN [Rasbash et al (2004)] statistical software is used for all of the following analyses.

## 4.2 Descriptive test of heterogeneity

Before considering the model for comparing more than two groups, we conduct a descriptive analysis.To obtain the mean of child weight for each of the 11 regions in the sample.

Table 4.3: Mean and standard deviation of low birth weight of different regions of Ethiopia.

| Region | Tigray | Affar | Amhara | Oromiya | Somali |
|---|---|---|---|---|---|
| N | 365 | 361 | 467 | 586 | 302 |
| Mean | 0.608 | 0.781 | 0.651 | 0.404 | 0.570 |
| Standard dev. | 0.489 | 0.414 | 0.477 | 0.491 | 0.496 |

| Region | B. gumuz | SNNP | Gambela | Harari | Addis A. | Dire D. | Total |
|---|---|---|---|---|---|---|---|
| N | 322 | 493 | 315 | 204 | 87 | 213 | 3715 |
| Mean | 0.45 | 0.448 | 0.556 | 0.442 | 0.414 | 0.460 | 0.532 |
| St. deviat. | 0.498 | 0.498 | 0.498 | 0.495 | 0.495 | 0.500 | 0.499 |

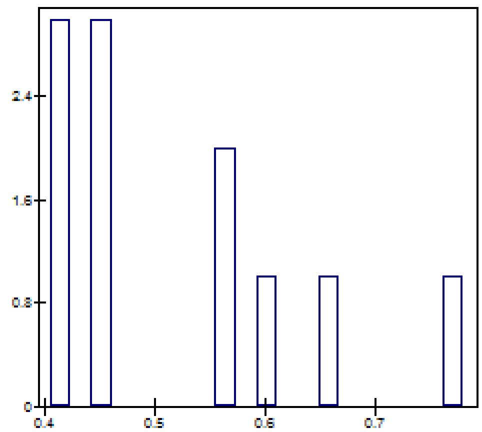



Figure 4.1: Histogram of the region means of child weight.

Table 4.4: ANOVA table for between and within regional variability.

| variation | Df | Ss | Ms | F |
|---|---|---|---|---|
| Between region | 10 | 51.686 | 5.17 | 21.93 |
| Within region | 3704 | 873.16 | 0.236 | |
| Total | 3714 | 924.84 | 0.249 | |

From the table 4.3 we can conclude that there is clear low birth weight variability between and within region. It is helpful to display the distribution graphically using a histogram. From the histogram, we see that there is a large amount of variation in the mean of child weight across regions.Since the width of the bars are different in length. From the table 4.4 we can conclude that there is significant variability of low birth weight babies between regions [$F_{cal} = 21.93 > F_{(10,3704)} = 0.394\ at\ 5\%\ significance\ level$]. The within and between regional variability is 0.49 and 0.122 in standard deviation respectively.

### 4.3 Intercept Only Multilevel Logistic Model [Empty model]

We first estimated a model with no predictors i.e. an intercept-only model that predicts the probability of low birth weight infants. The estimates of parameters and standard errors are presented in Table 4.4.The likelihood estimate from the standard logit model of the odds ratio of low birth weight to high birth weight is 1.14, which is the same as the sample ratio of 1978 low birth weight to 1737 high birth weight. It is in fact odds-ratio when no predictors have been considered in the model. In comparison, the same ratio is estimated to be 1.113, 1.13, 1.12, 1.13 and 1.132 from the multilevel model by the MQL-1, MQL-2, PQL-1, PQL-2 and MCMC methods respectively. Compared to the odds-ratios obtained by all multilevel methods the standard logistic model odds-ratio has overestimated. It is observed that there is a significant difference between the standard logistic estimate and the multilevel logistic estimate. Significant difference is also seen between their standard errors. Therefore, failing to take into account the clustering within region (level 3) and household (level 2), the standard logistic model has overestimated the odds-ratio by about 21.5%, 16.07%, 8.33% and 10.17% when multilevel model by corresponding methods MQL-1, MQL-2, PQL-1 and PQL-2 has been applied (see Table 4.4). The random quantity including its standard error at household level is similar to all methods, but the region level random quantity is too high compared to all other methods. When multilevel effects are considered in the model the model estimate reflects the value closer to the real value. The above analysis also demonstrates large differences among the estimates from different estimation methods. For instance, the estimated variances of the random effect at the region level (level 3) are respectively 0.064, 0.064, 0.072, 0.078 and 0.097 for MQL-1, MQL-2, PQL-1, PQL-2, and MCMC (but differences between MQLs or PQLs are less). The differences among the estimated fixed effects (also the odds-ratios) are also large.

Table 4.5: Parameters and standard errors of an intercept-only logit model and an intercept-only multilevel model predicting the probability of low birth weight (S.E.s are placed in parentheses)

| Model effect | Standard logit | Multilevel models | | | | |
|---|---|---|---|---|---|---|
| | logit | MQL-1 | MQL-2 | PQL-1 | PQL-2 | MCMC |
| Fixed effect | | | | | | |
| Intercept | 0.130 | 0.107 | 0.120 | 0.112 | 0.118 | 0.124 |
| | (0.033) | (0.046) | (0.046) | (0.049) | (0.051) | (0.054) |
| Random effect | | | | | | |
| Intercept(level-2) | | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Random effect | | | | | | |
| Intercept(level-3) | | 0.438 | 0.439 | 0.52 | 0.597 | 0.674 |
| | | (0.064) | (0.064) | (0.072) | (0.078) | (0.097) |
| -2logL | 5216.883 | | | | | |
| Deviance | | | | | | 4634.253 |
| N | 3715 | 3715 | 3715 | 3715 | 3715 | 3715 |

The parameters under random effect in Table 4.5 are the estimated variances of the random intercepts at both levels (level 2 & 3) for fitting a three level intercept-only model. In this three level intercept-only model to understand the random effect, we can imagine a unique effect for each region (level 3) and for each household (level 2) in addition to the fixed intercept of 0.118 (PQL-2 estimate), which is the average of all regions or all household. The addition of the household specific effects as well as region specific effects makes the model more accurate than the fixed intercept only model.

## 4.4 Random Intercept Multilevel Logistic Regression Model

Now in this univariate analysis we will find whether each of the explanatory variables is influencing the response 'LBW' and how much the estimates distorted from the actual when multilevel effect has not considered. The results of the univariate multilevel logistic models are presented in Table 4.6. For each multilevel model we present additional components, i.e. the household-level and region-level variance components and their standard errors in the table. The fifth and eighth column of Table 4.6 represents odds ratios of the standard logistic model and multilevel model respectively. In logistic regression, the odds of outcome for a non reference case in a predictor variable divided by the odds of outcome for a reference case for the same predictor variable does not depend on the other predictor variables. Thus, although odds ratios can be calculated from the log odds, odds ratios cannot be isolated in the multilevel model in a comparable manner to logistic regression. To correctly interpret the parameter estimates related to predictors in a multilevel model, it is more meaningful to state that the individual estimates increase or decrease the log odds of the outcome. Another possibility is to convert the log odds into probabilities. Instead in this study the odds ratios have been roughly compared each other. It is observed that there exist significant differences between the odds ratios of these two models for each of the explanatory variables. Also the odds ratios of the standard model have been underestimated in comparison with the multilevel model. The difference in the odds ratios estimated from a multilevel and standard model arises because of the addition of the random effects. These differences imply that a single-level model for this outcome variable is not appropriate.

Table 4.6 shows that whether the multiple births are found to be a significant predictor of low birth weight in Ethiopia. When these multilevel effects have not been taken into consideration, the odds ratios have been underestimated for a multiple birth. For instance, for a  multiple birth the odds ratio of single level model have been under estimated by 4.2%.The odds of low birth weight among  multiple birth is about 2.64 times higher than the odds among  single birth under the multilevel model whereas under the single level model the corresponding odds is 2.53 times higher.

Table 4.6 shows that sex of a child is found an important predictor variable. Male children are significantly less likely to have low birth weight than Female children. Multilevel analysis from Table 4.6 reveals that the odds of low birth weight among Female children is about 74% higher than the odds among male children. But from the single level analysis the figure is found 66%. Also there exist significant variation in the intercepts at region level (at 5%) and non-significant variation at household level (at 5%). The positive slope implies that probability of low birth weight is higher for female children than male children.

Table 4.6 reveals that wealth index (WI) or maternal economic status is another significant determinant of low birth weight. The probability of low birth weight is higher among the women who are from economically well off families. The multilevel analysis shows that the women from poorer, middle, richer and richest economic status have odds of low birth weight 27%, 26%, 52% and 54% lesser compared to the odds among poorest women. The corresponding figures under single level model are about 29%, 32%, 54% and 54%.The p-value (0.000) shows that the average response obtained from some region affected by this variable is significantly different from that of other region.

Results of Table 4.6 are related to current age of the women when we considered age as a continuous variable measured in years. The results show that maternal age is not significantly influencing low birth weight though there is a non- significant random effect at household level and significant (at 5%) random effect at region level. But when maternal age is considered as a categorical variable the univariate model shows that the low birth weight largely depends on a maternal age category. The positive slope for age group 20-39 implies that probability of low birth weight increases for this age group as compared to age group 10-19.The negative slope for age group 40-49 implies that probability of low birth weight decreases for this age group as compared to age group 10-19. For instance, for age category 20-39 and 40-49 the odds ratios of single level model have been underestimated by almost 2.3% and 1.3% respectively. The odds of low birth weight among women of age group 20-39 is about 1.33 times higher than the odds among women of age group 10-19 under the multilevel model whereas under the single level model the corresponding odds is 1.30 times higher. The multilevel analysis shows that the odd of low birth weight among women of age group 40-49 is 23% lesser compared to the odds among age group 10-19. The corresponding figure under single level model is about 24%.

There is a clear relationship of maternal ‘place of residence’ with low birth weight. The probability of low birth weight by the women in rural places is significantly more than that for the urban women. The odds of low birth weight among rural women are about 52% higher than the odds among rural women in multilevel analysis whereas in single level analysis the corresponding figure is 35%. The intercept of this variable is also found to have varied considerably between regions. This implies that in some regions the gap between the low birth weight of urban and rural women is higher while in some region the gap is lower.

Table 4.6: Parameters, standard errors and odds ratios of univariate single level logistic model and univariate multilevel model predicting the probability of low birth weight with random intercept and fixed slope using PQL-2 method (S.E.s are placed in parentheses).

| covariates | Single level model | | | Multilevel model | | | | |
|---|---|---|---|---|---|---|---|---|
| | Intercept | slope | ψ | Intercept | slope | ψ | household (level-2) | region (level-3) |
| 1.sex | | | | | | | | |
| Female | -0.116 * | 0.507* | 1.66 | -0.158* | 0.566* | 1.74 | 0.00 | 0.61* |
| | (0.046) | (0.066) | | (0.062) | (0 .072) | | (0.00) | (0.080) |
| 2.wealth index | | | | | | | | |
| Poorer | 0.505* | -0.349* | 0.71 | 0.486 * | -0.318 * | 0.73 | 0.00 | 0.512* |
| | (0.059) | (0.097) | | (0.077) | (0.108) | | (0.00) | (0.072) |
| Middle | (-) | -0.381* | 0.68 | (-) | -0.296 * | 0.74 | (-) | (-) |
| | (-) | (0.098) | | (-) | (0.112) | | (-) | (-) |
| Richer | (-) | -0.775* | 0.461 | (-) | -0.729 * | 0.48 | (-) | (-) |
| | (-) | (0.101) | | (-) | (0.117) | | (-) | (-) |
| Richest | (-) | -0.771* | 0.463 | (-) | -0.768* | 0.46 | (-) | (-) |
| | (-) | (0.104) | | (-) | (0.125) | | (-) | (-) |
| 3.Residence | | | | | | | | |
| Rural | -0.126* | 0.298* | 1.35 | -0.227* | 0.421* | 1.52 | 0.00 | 0.599* |
| | (0.088) | (0.095) | | (0.115) | (0.126) | | (0.00) | (0.078) |
| 4.Education level | | | | | | | | |
| Primary | 0.261* | -0.490* | 0.61 | 0.258* | -0.476 * | 0.62 | 0.00 | 0.560* |
| | (0.038) | (0.080) | | (0.055) | (0.088) | | (0.00) | (0.076) |
| Secondary | (-) | -0.677* | 0.32 | (-) | -0.666* | 0.51 | (-) | (-) |
| | (-) | (0.235) | | (-) | (0.256) | | (-) | (-) |
| Higher | (-) | -0.867* | 0.42 | (-) | -0.844 * | 0.43 | (-) | (-) |
| | (-) | (0.361) | | (-) | (0.392) | | (-) | (-) |
| 5.Anemia | | | | | | | | |
| Moderate | 0.887* | -0.465* | 0.63 | 0.831* | -0.437* | 0.65 | 0.00 | 0.577* |
| | (0.318) | (0.124) | | (0.345) | (0.122) | | (0.00) | (0.077) |
| Mild | (-) | -0.627* | 0.53 | (-) | -0.604* | 0.55 | (-) | (-) |
| | (-) | (0.301) | | (-) | (0.300) | | (-) | (-) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Not anemic | (-) | -0.830* | 0.44 | (-) | -0.773* | 0.46 | (-) | (-) |
| | (-) | (0.32) | | (-) | (0.346) | | (-) | (-) |
| 6.Multiple Birth | | | | | | | | |
| Multiple | 0.114* | 0.928* | 2.53 | 0.101 | 0.971* | 2.64 | 0.00 | 0.598* |
| | (0.033) | (0.276) | | (0.05) | (0.296) | | (0.00) | (0.078) |
| 7.Antenatal visits | | | | | | | | |
| No visits | -0.133* | 0.441* | 1.55 | -0.139* | 0.452* | 1.57 | 0.00 | 0.577* |
| | (0.052) | (0.067) | | (0.066) | (0.076) | | (0.00) | (0.077) |
| 8.Maternal weight | | | | | | | | |
| | 0.127* | -0.001* | 1.00 | 0.127* | -0.001 | 1.00 | 0.00 | 0.427* |
| 9.Maternal height | (0.033) | (0.00) | | (0.033) | (0.00) | | (0.00) | (0.063) |
| | 0.130* | 0.00 | 1.00 | 0.118 * | 0.00 | 1.00 | 0.000 | 0.598* |
| | (0.033) | (0.00) | | (0.051) | (0.00) | | (0.00) | (0.078) |
| 10. Maternal age | | | | | | | | |
| | 0.220 | -0.003 | 1.00 | 0.310 | -0.006 | 0.99 | 0.000 | 0.601* |
| | (0.157) | (0.005) | | (0.175) | (0.005) | | (0.000) | (0.078) |
| Age category | | | | | | | | |
| 20-39 | 0.261 | 0.264 * | 1.30 | 0.281 | 0.282* | 1.33 | 0.000 | 0.521* |
| | (0.170) | (0.047) | | (0.110) | (0.005) | | (0.000) | (0.072) |
| 40-49 | (-) | -0.278* | 0.76 | (-) | -0.297* | 0.74 | (-) | (-) |
| | (-) | (0.006) | | (-) | (0.020) | | (-) | (-) |

Note: The symbol * indicate that the estimate is significant at 0.05. Reference categories are:"single" for a multiple Birth , "male" for sex of a child, "poorest" for wealth index, "urban" for place of residence, "sever" for anemia level, "no education" for education level, "10-19" for mother's age category and "visits" for antenatal visits.

Education seems to be another influential factor of low birth weight. Among those women who have primary, secondary and higher education the respective odds of low birth weight is about 38%, 49% and 57% lesser compared to the odds of low birth weight among women without education for the multilevel model. For the single level model the corresponding odds ratios are 39%, 68%, and 58%.The estimates also differ significantly between multilevel and single level model. The coefficient of primary education is highly significant with LBW for multilevel model and for single level it is significant at 5%.Therefore, multilevel effect plays a vital role in measuring the true effect by the variable or by their categories. This factor also caused significant variation in the mean effect in different regions. The odds of low birth weight decrease with the level of education increases for both multilevel and single level models as compared to no education category. That means the probability of low birth weight by women with no education is significantly higher than that for educated women. The negative slope for education categories implies that low birth weight decreases for educated women than no educated women.

The results of table 4.6 shows that anemia level of the women is significantly influencing low birth weight though there is insignificant mean random effect at household level, but significant (5%) random effect at region level. But when anemia level is considered as categorical variable the univariate model shows that the low birth weight largely depends on woman's anemic category. The odds of low birth weight decreases with the level of anemia, for both multilevel and single level model as compared to sever anemic category. Among those women who have moderate, mild and non anemic categories the odds of low birth weight is about 35%, 45%, and 54% lesser as compared to the odds of low birth weight among women with sever anemic category for the multilevel model. For single level model the corresponding odds ratio are 37%, 47% and 56%. The negative slope implies that the probability of low birth weight decreases with the level of anemia.

Antenatal care is found an important predictor variable. No antenatal visit women are significantly higher LBW than antenatal visit women. Multilevel analysis from Table 4.6 reveals that the odds of low birth weight among no antenatal visit women is about 57% higher than the odds among antenatal visit women. But from the single level analysis the figure is found only 55%. Also there exists significant variation in the intercepts for regions (at 5%).

The result of table 4.6 shows that maternal weight is significantly influencing low birth weight though there is a significant mean random effect at region level, but not at household level. The result shows that height is not significantly influencing low birth weight though there is significant mean random effect at region level, but not at household level.

## 4.5 Random Slope Multilevel Univariate Model

Earlier we considered fitting intercept-only model with random intercept (table 4.5) and then multilevel models with random intercept for each predictor separately. Now this univariate multilevel model with random intercept can be expanded into an univariate multilevel model with random slope. The variance components models (table 4.6) that we have just worked with assume that the only variation between households or regions is in their intercepts i.e. the slope was fixed for all households and for all regions.

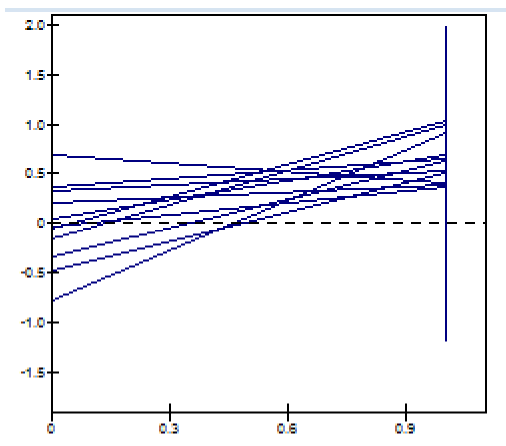



Figure 4.2: Region level (level-3) predicted line fitted by univariate model with random intercept and slope for Sex. From top to bottom the lines corresponding to Tigray, Amhara, Somali, Harari, Beneshangle Gumuzi, SNNP, Gambela, Afar, Addis Ababa, Dire Dawa,and oromiya regions respectively.

There is the possibility that the regions have different slopes i.e. the probability of low birth weight varies across regions. This implies that the coefficient of the explanatory variable sex will vary from region to region. In the univariate analysis we can achieve this by fitting a random effects model for each of the explanatory variables. We have found that the sex of child variable has significant random effects across regions ($\hat{\sigma}_{v1}^2 = 0.032\ with\ SE = 0.000$), not across households (because$\hat{\sigma}_{u1}^2 = 0.000\ with\ SE = 0.000$). That is, the model allows the difference between male and female children vary across regions. Figure 4.2 shows that the intercept and slopes of Sex vary across the eleven regions. The top line in the graph represents the predicted line for ‘Tigray’ region. Tigray region has highest intercept while Oromiya region (bottom line in Figure 4.2) has highest slope value. The three level random model for sex of child can be written as below,

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \beta_{0jk} + \beta_{1jk}sex_{ijk} \quad (4.1)$$

Where, $\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$ and $\beta_{1jk} = \beta_1 + v_{1k} + u_{1jk}$. The Sex is a binary variable with categories ‘Male’ and ‘Female’. Taking ‘Male’ as reference category.

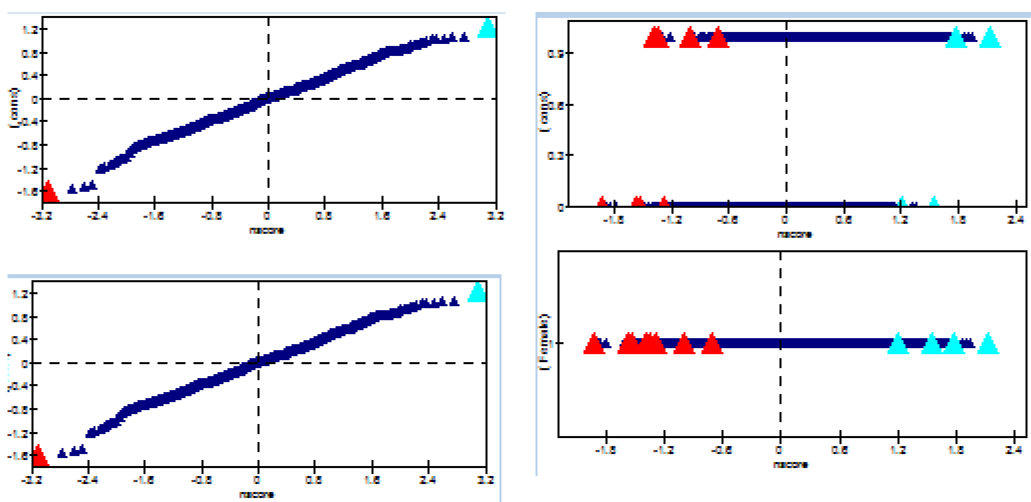
## 4.6 Multilevel Multivariate Logistic Model

The immediate univariate analysis plays a role first to detect whether the factors individually affect the low birth weight of babies from EDHS 2011 survey and second to what extent the effect is existed. The relationship between low birth weight babies and each of our selected predictors has been found statistically significant in both single and multilevel univariate analysis. Now the multivariate logistic model is followed with those significant factors to assess their simultaneous effect on low birth weight babies. To predict the probability that a children aged 0-59 weeks will have low birth weight, we need to know their sex, multiple birth, maternal income level, maternal education level, her residence, her age, anemia level and antenatal care. In both single and multilevel analysis multivariate regression approach it follows that multiple birth, maternal education level, maternal residence, maternal age and maternal anemia level do not have a statistically significant effect over the response when other factors have been considered in the model. Although this variable had played a significant role in univariate modeling, it is not statistically significant in the multivariate analysis, probably due to the presence of multi co linearity with some other model variables like sex of a child, wealth index of mothers and no antenatal visit.In the beginning of fitting multilevel multivariate model we considered a full random effects model i.e. we allowed random effects for intercept and other variables in the model. Only random effects for Sex of child and intercept have been found statistically significant. In earlier univariate analysis random effect of Sex of a child was also found significant. We describe the multilevel model as follows:

$$\ln\left[\frac{p_{ijk}}{1-p_{ijk}}\right] = \beta_{0jk} + \beta_{1jk}sex_{ijk} + \beta_2 wealth_{ijk} + \beta_3 no\ antenatal\ visit_{ijk} \quad (4.3)$$

Where, $\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$, and $\beta_{1jk} = \beta_1 + v_{1k} + u_{1jk}$

In fiting of this model we also went through data exploration or a diagnostics technique which is still a little explored area of multilevel modeling. In data structures of increasing complexity like our three level modeling, the concept of an outlier becomes less clear-cut. We may wish to know at what level(s) a particular response is outlying, and in respect to which explanatory variable(s). Our data is of a 3-level structure with children nested within households, and households nested within regions. Now question is whether children, households or regions may be considered as being outliers at their respective levels in the model. Suppose, for example, that at the region level a particular region is found to be a discordant outlier; we will need to ascertain whether it is discordant due to a systematic deference affecting all the children measured within that region, or because one or two children are responsible for the discrepancy. At the children level, an individual may be outlying with respect to the overall relationships found across all regions, or be unusual only in the context of his or her particular region. Our aim is to show diagnostics at region level.



(a) (b )

Figure 4.3: (a) Normal probability plot of region level residuals for intercept(top) and Sex (bottom), (b) Region level (level-3) predicted plot for intercept (top) and Sex (bottom).

The model 4.3 has two random components: intercept and sex or female. This model fitted well with these two random components. Figure 4.3 shows the pattern of their residual (standardize or student zed) and predicted values at region level. Normal probability plot (a) for both intercept and sex are almost straight despite of somewhat distorted at two ends. We can see, there are two regions, one at the bottom left (red triangle) and other at the top right (sky triangle) of the plots who have particularly high negative and positive residuals respectively. The household number with highest negative residual is 538 which is nested within oromiya region. On the other hand the number of household with positive residual is 33 which is nested within Affar region. Predicted plot (b) for both intercept and sex at region level shows that the predicted values of household 538 and 33 for intercept and household 570 and 33 for sex are not too far from the standard range.

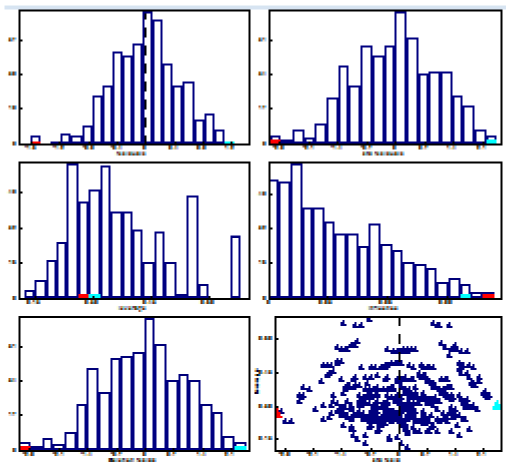


Figure 4.4: Diagnostic plotting: residual, student zed residual, leverage, influence, Deletion residuals for intercept at region level for multilevel multivariate logistic model.

Though the predicted values of these two regions are within accepted limit there might be their significant influence on the prediction. Figure 4.4 gives some diagnostic plots of understanding outliers, leverages and residuals. The figure represents six plots of diagnostic measures associated with the intercept at the region level. Oromiya region and Afar which we have previously chosen to high light, are shown red and sky on all six diagrams. The plot shows that Afar region has highest intercept residual, standardized residual and deletion residual. Oromiya region has also significantly higher influence value. On the contrary household 33 has highest intercept residual, standardized residual, leverage and Deletion residual but does not have a particularly high influence value. Though we have not shown here, we have got almost similar results, from the measures and plots for the slopes at the region level associated with the explanatory variable sex. Therefore, household 33 and 570 can be treated as extreme outliers since they have unusual residual, leverage and influence. We finally fit the model 4.3 omitting these two households which leave total number of households 887 for analysis. The last two columns of table 4.7 represent respectively the difference in odds ratio between single and multilevel multivariate models and percentage of under or over estimation of odds ratio by single level multilevel modeling.

Table 4.7: Parameters, standard errors and odds ratios of single level multivariate logistic model and multilevel multivariate model predicting the probability of low birth weight with random intercept, random slope for Sex and fixed slope for others using PQL-2 method (S.E.s are placed in parentheses).

| Parameter | Single level model Estimate | $\hat{\psi}_1$ | Multilevel model Estimate | $\hat{\psi}_2$ | $\hat{\psi}_1$-$\hat{\psi}_2$ | over/under Estimation (%) |
|---|---|---|---|---|---|---|
| Fixed parameter | | | | | | |
| Intercept | 0.259(0.067)* | 1.3 | 0.214(0.085)* | 1.24 | 0.06 | 4.6 |
| Sex | | | | | | |
| Female | 0.532(0.068)* | 1.7 | 0.578(0.072)* | 1.78 | -0.08 | 4.7 |
| Wealth Index | | | | | | |
| Poorer | -0.326(0.098)* | 0.72 | -0.292(0.109)* | 0.75 | -0.03 | 4.2 |
| Middle | -0.342(0.100)* | 0.71 | -0.253(0.114)* | 0.78 | -0.07 | 10.0 |
| Richer | -0.730(0.103)* | 0.48 | -0.685(0.119)* | 0.50 | -0.02 | 4.2 |
| Richest | -0.656(0.111)* | 0.52 | -0.640(0.131)* | 0.53 | -0.01 | 2.0 |
| Antenatal visits | | | | | | |
| No visits | 0.296(0.072) * | 1.34 | 0.344(0.080)* | 1.41 | -0.07 | 5.2 |
| Random parameter | | | | | | |
| $\sigma_{u0}^2$(Intercept) | | | 0.00(0.00) | | | |
| $\sigma_{u1}^2$(Sex) | | | 0.00(0.00) | | | |
| $\sigma_{u01}$ | | | 0.00(0.00) | | | |
| $\sigma_{v0}^2$(Intercept) | | | 0.523(0.074)* | | | |
| $\sigma_{v1}^2$(Sex) | | | 0.420(0.01)* | | | |
| $\sigma_{v01}$ | | | -0.000(0.00) | | | |

**Note: The symbol $*$ indicate that the estimate is significant at 0.05.Refrence categories are: "Male" for Sex, "Poorest" for Wealth Index and "visit" for antenatal care.**

.

The multivariate model shows that the probability of low birth weight decreases significantly with wealth index, adjusting for the effect of other predictors. When all other predictors are fixed in multilevel analysis the probability of low birth weight decreases 25%, 22%, 50% and 47% as WI category poorer, middle, richer and richest, but for single level analysis the percentage is about 28%, 29%, 52% and 48% respectively. The odds ratio under single level model for middle income group is under estimated (10%) compared to multilevel estimates. That is, the multilevel effect is observed notably for predictor WI for category middle.

In the multivariate framework variables sex of child and no antenatal visits have been found significantly associated with low birth weight. Also the multilevel impact on each of these variables is very high. . The odds ratios for variable sex of child and no antenatal visits, from standard logit model, have been underestimated by 4.7% and 5.2%, respectively. Thus it is evident that if multilevel effect is not taken into account in multivariate modeling the estimates would be under estimated considerably. These results imply that single-level multivariate model for this outcome variable is not appropriate. The multilevel multivariate model has also revealed that there exist variations in the mean effect of the predictors over the response variable low birth weight in Ethiopia. The variation is significant ($p < 0.000$) at all levels of the hierarchy (lower, middle, and higher).In addition to the fixed effect the intercept and sex of child has very strong significant random effects at region level, but not at household level. This indicates that variation among the mean effects obtained from 887 households and 11 regions due to the linear combination of the selected variables are significant.

## 4.7 Discussions

The purpose of multilevel model is to evaluate the household and regional variability of low birth weight in Ethiopia. Three models were fitted one with intercept only model used to check the mean effect of response variable (LB W) without predictor variable among households and regions. The second model is called random intercept model is used to find out whether each of the explanatory variable is influencing the response (LBW) and how much the estimate distorted from the standard logistic regression model when multilevel effect has not considered. The third model is used to check there is the possibility that the regions have different slopes and also the households have different slopes i. e the probability of low birth weight varies across both households and regions. Our study found that for such hierarchical structured data the multilevel effects have been found significant and have to be taken into consideration in logistic regression modeling. As a result, this multilevel analysis enables the proper investigation of the effects of all independent variables measured at different levels (households and regions) on the response variable low birth weight.

The univariate multilevel analysis (Table 4.6) of this study revealed that each of the predictors over low birth weight with multiple birth, sex of child, wealth index of mothers, education level of mothers, maternal place of residence, maternal anemia level, maternal no antenatal visit varied much significantly ($p < 0.000$), this result agrees with the result of [Dharma lingam, et al, 2010, khatun,S., &Rah man, M. 2008, Ipadeola, O.B., et al, 2013.] . And the predictor 'sex of child' has significant random effect in region level but not at household level. Similar type of significance results have been found in multivariate multilevel analysis for sex of child, maternal income level and no antenatal visit (Table 4.7). Mean effects of the combination of the predictor's sex of child, wealth index and no antenatal visit are varied significantly in region level but not in household level and random effect of sex of child varied considerably in region level. Thus through this study it is now noticed that different regions have significantly different mean effects as well as the effect for sex of child is different in female and male infants across the regions, which had remained unknown in the studies with the single level modeling approach done so far.

The results  of  table 4.6 shows the negative association between wealth status of maternal and LBW which agree with study done in England by Smith, G. C., *et al* (2010) and in Ghana Charles *et al* (2011). The study shows that the odds of maternal bearing child with LBW are consistently decreased as the mother wealth status increased. One of the most predominant causes of low birth weight is the maternal age. The chance of having LBW baby is higher among young mothers of age 10-19 as compared to age group 40-49. This is similar with finding of Kamaladoss *et al*, 2013.The positive association between multiple birth and LBW agree with study done by Ipadeola, o. B, et al, 2013. The study shows that  the odds of multiple birth is 2.64 times higher than single birth for multilevel model and the corresponding figure is 2.53 times higher  for single level model. There is also a significant association between LBW and maternal anemia. According to this study, maternal anemia increased the risk of having a LBW baby. The findings of this study are similar to a study done in Turkey by Chuku, S. N., 2013. In agreement with previous studies, maternal education emerged as a strong determinant for LBW. Women with ‘no education’ had the greatest odds of giving birth to an infant with LBW. This finding is similar with some other studies such as, Karim E, *et al*.2012. This study showed the positive effect of number of no antenatal visit on LBW. The positive slope for no antenatal visit implies that the probability of LBW increases for this group than antenatal visit women. Those mothers received antenatal care gave birth to higher birth weight babies in comparison to mothers who do not received antenatal care visit. The other studies also found similar result. Naher N, *et al*,. 2012.

## Chapter 5

## Conclusions and Recommendations

### 5.1 Conclusions

For this study three level multilevel logistic regression model families, empty model, random intercept model and random slope model, have been employed for the analysis of effects of covariates on the response variable (LBW) and, we conclude that multilevel multivariate logistic regression model we considered a random effect for intercept and sex of child and fixed effect for maternal wealth index and maternal no antenatal visit is best fit for this data. Random effects for sex of child and intercept have been found to be statistically significant at region level. This study suggests that sex of child, maternal wealth index and maternal no antenatal visit have been found simultaneously statistically significant. But univariate analysis shows that sex of child, maternal wealth index, maternal place of residence, maternal education level, maternal anemia level, multiple birth, maternal age and maternal weigh have been found statistically significant and are varies across region.

### 5.2 Recommendations

With regard to decreasing low birth weight, there are a number of strategic implications that flow from these findings. It is necessary to create greater awareness among 20-39 aged women of the issues which are found to affect low birth weight significantly. All opportunities, namely, the school system, youth association, ministry of health and its hierarchal offices should be considered to educate 10-49 women aged about nutrition, iron and vitamins supplementation during pregnancy. Women should pay attention on practicing eating nutrition supplemented foods in their early reproductive lives. Poor women have high rate of low birth weight compared to middle and rich women. One possible reason could be that poor women in Ethiopia enjoy less nutrition developed food. Thus instead of widening the ongoing nutrition facilities to all regions equally, it is useful to give special attention to those regions where the low birth weight remained high.

## References

Agresti, A., (2002).Categorical Data Analysis, Second Edition

Agresti, A., (2007).An Introduction to Categorical Data Analysis, 2nd Ed, Wiley Inc

Ahmed, F. U., & Das, A. M. (2009). Beneficial effects of Three ANC visits might be the divergent point in lowering low birth weight babies. Bangladesh. *Integration*, (33), 50-3.

Amory JH, Adams KM, Lin MT, Hansen JA, Eschenbach DA, Hitti J. (2013) anemia and prenatal outcome. *Prenatal Journal*, 15(3): 127-130.

Anderson R. and Bergstrom S. (2013) 'Maternal nutrition and socioeconomic status as determinants of birth weight in chronically-malnourished African women', *Tropical Medicine and International Health* 2(11): 1080-87

Andersson SW, Niklasson A, Lapidus L, (2013) Sociodemo- graphic characteristics influencing birth outcome in Sweden,. J Epidemiology Community Health 2000; 54:269-78.

Antonio, K. & Beirlant, J., (2011).Actuarial Statistics with Generalized Linear Mixed Models, University Center for Statistics, Belgium.

Barbieri, M. A., Silva, A. A., Bettiol, H., & Gomes, U. A. (2014) Risk factors for the increasing trend in low birth weight among live births born by vaginal delivery, Brazil. *Revista de Saúde Pública*, *34*(6), 596-602.

Belsley D. A., E. Kuh and R. E. Welsch. (1980) Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

Berihun, Zeleke, Megabiaw, Meseret Zelalem, and Nuru Mohammed. (2012) "Incidence and correlates of low birth weight at a referral hospital in Northwest Ethiopia." *Pan African Medical Journal* 12(1)

Bradley, R. H., & Corwyn, R. F. (2011). Socioeconomic status and child development. *Annual review of psychology*, *53*(1), 371-399.

Brawarsky, P., Stotland, N. E., Jackson, R. A., Fuentes-Afflick, E., Escobar, G. J., Rubashkin, N., & Haas, J. S. (2012). Pre-pregnancy and pregnancy-related factors and the risk of excessive or inadequate gestational weight gain. *International Journal of Gynecology & Obstetrics*, *91*(2), 125-131.

Central Statistical Agency, (2011). Ethiopian Demographic and Health Survey Addis Ababa Ethiopia. Channon, A. A. (2011). Can mothers judge the size of their newborn?

Assessing the determinants of a mother's perception of a baby's size at birth. *Journal of biosocial science*, *43*(05), 555-573. Chuku, S. N. (2013). *Low Birth Weight in Nigeria: Does Antenatal Care Matter?*

Cogswell, M. E., & Yip, R. (2011, June). The influence of fetal and maternal factors on the distribution of birthweight. In *Seminars in perinatology* (Vol. 19, No. 3, pp. 222-240).

Cook R. D. and Weisberg S. (1982). Residuals and influence in regression. New York:

Chapman and Hall. Cosmas, G., (2011).Socio-Demographic determinants of anemia among children aged 0-59 months in mainland Tanzania.

DaVanzo, J., J.P Habcht and W.P. Butz (2005) 'Assessing socioeconomic correlates of birth weight in Peninsular Malaysia: Ethnic differences and changes over time', *Social Science and Medicine* 18(5):387-404.

Davidson, A. C. and Snell, E. J. (1991). Residuals and diagnostics.

Dharmalingam, A., Navaneetham, K., & Krishnakumar, C. S. (2010). Nutritional status of mothers and low birth weight in India. *Maternal and child health journal*,*14*(2), 290-298.

Dičkutė, J., Padaiga, Ž., Grabauskas, V., Gaižauskienė, A., Basys, V., & Obelenis, V. (2012). Do maternal social factors, health behavior and working conditions during pregnancy increase the risk of low birth weight in Lithuania?. *Medicina*, *38*(3), 321-332.

Dubois, L., & Girard, M. (2006). Determinants of birthweight inequalities: Population-based study. *Pediatrics International*, *48*(5), 470-478.

Eisner, V, J.V. Brasie, M.W Pratt and A.C Hexter (2013) 'The risk of low birth weight', *American Journal of Public Health* 69(9):887-93.

Ester, W. A., & Hokken-Koelega, A. C. S. (2008). Polymorphisms in the IGF1 and IGF1R genes and children born small for gestational age: results of large population studies. *Best Practice & Research Clinical Endocrinology & Metabolism*, *22*(3), 415-431.

Ethiopia Demographic and Health Survey 2005 report Central Statistical Agency Addis Ababa, Ethiopia Faraway, J., (2006).Extending the Linear Model with R, Generalized Linear, Mixed Effects, and Nonparametric Regression Models, *Boca Raton London New York.*

Halbreich, U. (2011). The association between pregnancy processes, preterm delivery, low birth weight, and postpartum depressions—the need for interdisciplinary integration. *American journal of obstetrics and gynecology*, *193*(4), 1312- 1322.

Hirve SS, Ganatra BR. (2008) Determinants of low birth weight: a community based prospective cohort study.

Hosmer DW. And Lemeshow S. (2000). Applied logistic regression (2nd ed.). New York: Wiley & Sons. Ipadeola, O. B., Adebayo, S. B., Anyanti, J., & Jolayemi, E. T. (2013). Poverty levels and maternal nutritional status as determinants of weight at birth: An ordinal logistic regression approach. *International Journal of Statistics and Applications*, *3*(3), 50-58

Jolly, M., Sebire, N., Harris, J., Robinson, S., & Regan, L. (2011). The risks associated with low birth weight. *Human reproduction*, *15*(11), 2433-2437.

Kamaladoss T, Abel R, Sampathkumar V. (2013). Epidemiological correlates of low birth weight in rural Tamil Nadu. *Ind J Paed*; 59:209-304.

Karim E, Mascie-Taylor CG. (2012) The association between birth weight, sociodemographic variables and maternal anthropometry in an urban sample from Dhaka, Bangladesh. Ann Hum Biol. 1997; 24: 387- 401.

Khatun, S., & Rahman, M. (2008). Socio-economic determinants of low birth weight in Bangladesh: a multivariate approach. *Bangladesh Medical Research Council Bulletin*, *34*(3), 81-86.

Kramer, M (2004) 'Socioeconomic determinants of intrauterine growth retardation' *European of Clinical Nutrition* 52(S1):S21-S28

Kramer, M. (1998) 'Determinants of Low birth Weight: methodological assessment and metaanalysis', Bulletin of the World Health Organization 65:663-73.

Magadi, M., I. Diamond, N. Madise (2013) 'Individual and community-level factors associated with premature births, size of baby at birth and caesarean section.

McCullagh, P., & Nelder, J. (1989).Generalized Linear Models, 2ndedition, London Chapman and Hall 120-125

McCulloch, E., (1997).An Introduction to Generalized Linear Mixed Models, Biometrics Unit, and Statistics Center Cornell University.

Molenberghs, G., & Verbeke, G., (2005).Models for Discrete Longitudinal Data. Library of Congress

Mwabu, G. (2011) determinants of birth weight in Kenya.

Myers, H., Montgomery, C., Vining, G. &Robinson, J., (2010).Generalized Linear Models, With Applications in Engineering and the Sciences, Second Edition.

Naher N, Afroza S, Hossain M. (2010). Incidence of LBW in three selected communities of Bangladesh. Bangladesh Med Res Counc Bull. 24(2): 49-54.

Nair N, Rao RS, Chandrashekar S, Acharya D, Bhat HB.,( 2012). Socio-demographic and maternal determinants of low birth weight: A multivariate approach. Indian J Pediatr. 67(1): 914.

Olowonyo, T., S. Oshin, I. Obasanjo-Bello (2006) 'Some factors associated with low birth weight in Nigeria', Nigerian Medical Practitioner 49(6):154-7

Pan, W., (2001).Akaike's information criterion in generalized estimating equations, Biometrics, 57, 120-127

Pojda J. and Kelly L. (2000) 'Low Birth Weight' ACC/SCN Nutrition Policy Paper. A report based on the International Low Birth Symposium and Workshop held on 14-17 June 1999 at The International Centre for Diarrhoeal Disease

Preisser, J.S., Qaqish, B.F., (1996). Deletion diagnostics for generalised estimating equations. Biometrika 83, 551–562.

Rondo, P. H. C., Ferreira, R. F., Nogueira, F., Ribeiro, M. C. N., Lobert, H., & Artes, R. (2011). Maternal psychological stress and distress as predictors of low birth weight, prematurity and intrauterine growth retardation. European Journal of Clinical Nutrition, 57(2), 266-272.

Saurel-Cubizolles, M. J., Subtil, D., & Kaminski, M. (2013). Is preterm delivery still related to physical working conditions in pregnancy?. Journal of epidemiology and community health, 45(1), 29-34.

Sexton, M and J.R. Hebel (2007) 'A clinical trial of change in maternal smoking and its effect on birth weight', Journal of the American Medical Association 251:911-15

Siega-Riz, A. M., Adair, L. S., & Hobel, C. J. (2013). Maternal underweight status and inadequate rate of weight gain during the third trimester of pregnancy increases the risk of preterm delivery. Journal of Nutrition, 126(1), 146-153.

Silva, A. A., Barbieri, M. A., Gomes, U. A., & Bettiol, H. (2014). Trends in low birth weight: a comparison of two birth cohorts separated by a 15-year interval in Ribeirao Preto, Brazil. Bulletin of the World Health Organization, 76(1), 73.

Siza, J. E. (2008). Risk factors associated with low birth weight of neonates among pregnant women attending a referral hospital in northern Tanzania. Tanzania journal of health research, 10(1), 1-8

Smith, G. C., Smith, M. F., McNay, M. B., & Fleming, J. E. (2010). First-trimester growth and the risk of low birth weight. New England Journal of Medicine, 339(25), 1817-1822.

Som S Jr, Pal M, Adak DK, Gharami AK, Bharati S, Bharati P. (2012) Effect of socioeconomic and biological variables on birth weight in Madhya Pradesh. Malays J Nutr 10:159-71.

Tema T. (2006). Prevalence and Determinants of Low Birth Weight in Jimma Zone, Southwest Ethiopia. East African Medical Journal. Volume 83, pp.45-51

Tuntiseranee P, Olsen J, Chongsuvivatwong V, Limbutara S. (2013). Socioeconomic andwork related determinants of pregnancy outcome in Southern Thailand. J Epidemiol Community He- alth; Vol.53, pp.624-9

UNICEF and WHO (2004) 'Low birth weights: Country, Regional and Global estimates. UNICEF, Editorial and publication section, NY, USA.

Wardlaw, Tessa M., ed. (2004). Low Birthweight: Country, regional and global estimates.

Wedderburn, R.W.M. (1974).Quasi-Likelihood Functions, Generalized Linear Models and the Gauss Newton Method," Biometrika, Vol. 61, pp.439-447.

Zeger, S. & Liang K., (1986). Longitudinal data analysis using generalized linear models.Biometrics, Vol.73, pp.13-22

Zorn, C. J. (2001). Generalized estimating equation models for correlated data: A review with applications. American Journal of Political Science, 470-490.

## Appendix Table

Table 4.1 percentage of low birth weight (size of child) by predictors (covariates) results are based on our study sample=3715)

| Covariates | Measurement range | Birth weight | |
|---|---|---|---|
| | | Low | High |
| Sex | 0-1 | | |
| Male | | 47.1 | 52.9 |
| Female | | 59.6 | 40.4 |
| Wealth Index | 0-4 | | |
| Poorest | | 62.4 | 37.6 |
| Poorer | | 53.9 | 46.1 |
| Middle | | 53.1 | 46.9 |
| Richer | | 43.3 | 56.7 |
| Richest | | 43.4 | 56.6 |
| Place of residence | 0-1 | | |
| Urban | | 46.8 | 53.2 |
| Rural | | 54.3 | 45.7 |
| Education level | 0-3 | | |
| No education | | 56.5 | 43.5 |
| Primary | | 44.3 | 55.7 |
| Secondary | | 39.7 | 60.3 |
| Higher | | 35.3 | 64.7 |
| Anemia level | 0-3 | | |
| Sever | | 67.2 | 32.8 |
| Moderate | | 60.0 | 40.0 |
| Mild | | 56.4 | 43.6 |
| Not anemic | | 51.4 | 48.6 |

Table 4.2: Test of association: chi-square tests of independence between explanatory and dependent variable

| Independent variable | $x^2$-significance | Independent variable | $x^2$-significance |
|---|---|---|---|
| Multiple birth | $x_2^2$=12.959 (p<0.002) | Maternal Education level | $x_3^2$=48.411 (p<0.000) |
| Sex of child | $x_1^2$=58.657 (p<0.000) | Maternal anemia level | $x_3^2$=17.282 (p<0.0.01) |
| Maternal income | $x_4^2$=86.915 (p<0.000) | Preceding birth interval | $x_{139}^2$=142.06 (p<0.412) |
| Maternal age | $x_{33}^2$=27.795 (p<0.724) | Antenatal visit | $x_{17}^2$=91.459 (p<0.000) |
| Maternal residence | $x_1^2$=10.01 (p<0.002) | Marital status | $x_5^2$=6.667 (p<0.247) |
| Maternal weight | $x_{446}^2$=441.244 (p<0.295) | Maternal height | $x_{360}^2$=350.994 (p<0.623) |
| Region | $x_{10}^2$=207.617 (p<0.000) | Household | $x_{868}^2$=874.386 (p<0.433) |

Table 4.3: Mean and standard deviation of low birth weight of different regions of Ethiopia.

| Region | Tigray | Affar | Amhara | Oromiya | Somali |
|---|---|---|---|---|---|
| N | 365 | 361 | 467 | 586 | 302 |
| Mean | 0.608 | 0.781 | 0.651 | 0.404 | 0.570 |
| Standard dev. | 0.489 | 0.414 | 0.477 | 0.491 | 0.496 |

| Region | B. gumuz | SNNP | Gambela | Harari | Addis A. | Dire D. | Total |
|---|---|---|---|---|---|---|---|
| N | 322 | 493 | 315 | 204 | 87 | 213 | 3715 |
| Mean | 0.45 | 0.448 | 0.556 | 0.442 | 0.414 | 0.460 | 0.532 |
| St. deviat. | 0.498 | 0.498 | 0.498 | 0.495 | 0.495 | 0.500 | 0.499 |

Table 4.4: ANOVA table for between and within regional variability.

| variation | Df | Ss | Ms | F |
|---|---|---|---|---|
| Between region | 10 | 51.686 | 5.17 | 21.93 |
| Within region | 3704 | 873.16 | 0.236 | |
| Total | 3714 | 924.84 | 0.249 | |

Table 4.5: Parameters and standard errors of an intercept-only logit model and an intercept-only multilevel model predicting the probability of low birth weight (S.E.s are placed in parentheses)

| Model effect | Standard logit logit | Multilevel models MQL-1 | MQL-2 | PQL-1 | PQL-2 | MCMC |
|---|---|---|---|---|---|---|
| Fixed effect | | | | | | |
| Intercept | 0.130 | 0.107 | 0.120 | 0.112 | 0.118 | 0.124 |
| | (0.033) | (0.046) | (0.046) | (0.049) | (0.051) | (0.054) |
| Random effect | | | | | | |
| Intercept(level-2) | | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Random effect | | | | | | |
| Intercept(level-3) | | 0.438 | 0.439 | 0.52 | 0.597 | 0.674 |
| | | (0.064) | (0.064) | (0.072) | (0.078) | (0.097) |
| -2logL | 5216.883 | | | | | |
| Deviance | | | | | | 4634.253 |
| N | 3715 | 3715 | 3715 | 3715 | 3715 | 3715 |

Table 4.6: Parameters, standard errors and odds ratios of univariate single level logistic model and univariate multilevel model predicting the probability of low birth weight with random intercept and fixed slope using PQL-2 method (S.E.s are placed in parentheses).

| covariates | Single level model Intercept | slope | ψ | Multilevel model Intercept | slope | ψ | household (level-2) | region (level-3) |
|---|---|---|---|---|---|---|---|---|
| 1.sex | | | | | | | | |
| Female | -0.116 * | 0.507* | 1.66 | -0.158* | 0.566* | 1.74 | 0.00 | 0.61* |
| | (0.046) | (0.066) | | (0.062) | (0 .072) | | (0.00) | (0.080) |
| 2.wealth index | | | | | | | | |
| Poorer | 0.505* | -0.349* | 0.71 | 0.486 * | -0.318 * | 0.73 | 0.00 | 0.512* |
| | (0.059) | (0.097) | | (0.077) | (0.108) | | (0.00) | (0.072) |
| Middle | (-) | -0.381* | 0.68 | (-) | -0.296 * | 0.74 | (-) | (-) |
| | (-) | (0.098) | | (-) | (0.112) | | (-) | (-) |
| Richer | (-) | -0.775* | 0.461 | (-) | -0.729 * | 0.48 | (-) | (-) |
| | (-) | (0.101) | | (-) | (0.117) | | (-) | (-) |
| Richest | (-) | -0.771* | 0.463 | (-) | -0.768* | 0.46 | (-) | (-) |
| | (-) | (0.104) | | (-) | (0.125) | | (-) | (-) |
| 3.Residence | | | | | | | | |
| Rural | -0.126* | 0.298* | 1.35 | -0.227* | 0.421* | 1.52 | 0.00 | 0.599* |
| | (0.088) | (0.095) | | (0.115) | (0.126) | | (0.00) | (0.078) |
| 4.Education level | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Primary | 0.261* | -0.490* | 0.61 | 0.258* | -0.476 * | 0.62 | 0.00 | 0.560* |
| | (0.038) | (0.080) | | (0.055) | (0.088) | | (0.00) | (0.076) |
| Secondary | (-) | -0.677* | 0.32 | (-) | -0.666* | 0.51 | (-) | (-) |
| | (-) | (0.235) | | (-) | (0.256) | | (-) | (-) |
| Higher | (-) | -0.867* | 0.42 | (-) | -0.844 * | 0.43 | (-) | (-) |
| | (-) | (0.361) | | (-) | (0.392) | | (-) | (-) |
| 5.Anemia | | | | | | | | |
| Moderate | 0.887* | -0.465* | 0.63 | 0.831* | -0.437* | 0.65 | 0.00 | 0.577* |
| | (0.318) | (0.124) | | (0.345) | (0.122) | | (0.00) | (0.077) |
| Mild | (-) | -0.627* | 0.53 | (-) | -0.604* | 0.55 | (-) | (-) |
| | (-) | (0.301) | | (-) | (0.300) | | (-) | (-) |
| Not anemic | (-) | -0.830* | 0.44 | (-) | -0.773* | 0.46 | (-) | (-) |
| | (-) | (0.32) | | (-) | (0.346) | | (-) | (-) |
| 6.Multiple Birth | | | | | | | | |
| Multiple | 0.114* | 0.928* | 2.53 | 0.101 | 0.971* | 2.64 | 0.00 | 0.598* |
| | (0.033) | (0.276) | | (0.05) | (0.296) | | (0.00) | (0.078) |
| 7.Antenatal visits | | | | | | | | |
| No visits | -0.133* | 0.441* | 1.55 | -0.139* | 0.452* | 1.57 | 0.00 | 0.577* |
| | (0.052) | (0.067) | | (0.066) | (0.076) | | (0.00) | (0.077) |
| 8.Maternal weight | | | | | | | | |
| | 0.127* | -0.001* | 1.00 | 0.127* | -0.001 | 1.00 | 0.00 | 0.427* |
| 9.Maternal height | (0.033) | (0.00) | | (0.033) | (0.00) | | (0.00) | (0.063) |
| | 0.130* | 0.00 | 1.00 | 0.118 * | 0.00 | 1.00 | 0.000 | 0.598* |
| | (0.033) | (0.00) | | (0.051) | (0.00) | | (0.00) | (0.078) |
| 10. Maternal age | | | | | | | | |
| | 0.220 | -0.003 | 1.00 | 0.310 | -0.006 | 0.99 | 0.000 | 0.601* |
| | (0.157) | (0.005) | | (0.175) | (0.005) | | (0.000) | (0.078) |
| Age category | | | | | | | | |
| 20-39 | 0.261 | 0.264 * | 1.30 | 0.281 | 0.282* | 1.33 | 0.000 | 0.521* |
| | (0.170) | (0.047) | | (0.110) | (0.005) | | (0.000) | (0.072) |
| 40-49 | (-) | -0.278* | 0.76 | (-) | -0.297* | 0.74 | (-) | (-) |
| | (-) | (0.006) | | (-) | (0.020) | | (-) | (-) |

Table 4.7: Parameters, standard errors and odds ratios of single level multivariate logistic model and multilevel multivariate model predicting the probability of low birth weight with random intercept, random slope for Sex and fixed slope for others using PQL-2 method (S.E.s are placed in parentheses).

| Parameter | Single level model Estimate | $\hat{\psi}_1$ | Multilevel model Estimate | $\hat{\psi}_2$ | $\hat{\psi}_1$-$\hat{\psi}_2$ | over/under Estimation (%) |
|---|---|---|---|---|---|---|
| Fixed parameter | | | | | | |
| Intercept | 0.259(0.067)* | 1.3 | 0.214(0.085)* | 1.24 | 0.06 | 4.6 |
| Sex | | | | | | |
| Female | 0.532(0.068)* | 1.7 | 0.578(0.072)* | 1.78 | -0.08 | 4.7 |
| Wealth Index | | | | | | |
| Poorer | -0.326(0.098)* | 0.72 | -0.292(0.109)* | 0.75 | -0.03 | 4.2 |
| Middle | -0.342(0.100)* | 0.71 | -0.253(0.114)* | 0.78 | -0.07 | 10.0 |
| Richer | -0.730(0.103)* | 0.48 | -0.685(0.119)* | 0.50 | -0.02 | 4.2 |
| Richest | -0.656(0.111)* | 0.52 | -0.640(0.131)* | 0.53 | -0.01 | 2.0 |
| Antenatal visits | | | | | | |
| No visits | 0.296(0.072) * | 1.34 | 0.344(0.080)* | 1.41 | -0.07 | 5.2 |
| Random parameter | | | | | | |
| $\sigma^2_{u0}$(Intercept) | | | 0.00(0.00) | | | |
| $\sigma^2_{u1}$(Sex) | | | 0.00(0.00) | | | |
| $\sigma_{u01}$ | | | 0.00(0.00) | | | |
| $\sigma^2_{v0}$(Intercept) | | | 0.523(0.074)* | | | |
| $\sigma^2_{v1}$(Sex) | | | 0.420(0.01)* | | | |
| $\sigma_{v01}$ | | | -0.000(0.00) | | | |

**Note: The symbol $*$ indicate that the estimate is significant at 0.05.Refrence categories are: “Male” for Sex, “Poorest” for Wealth Index and “visit” for antenatal care.**

.

## Appendix figure

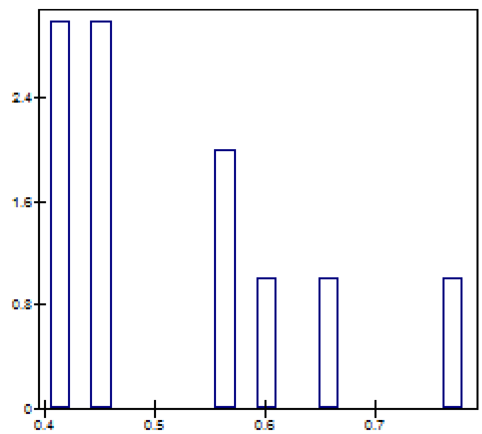



Figure 4.1: Histogram of the region means of child weight.

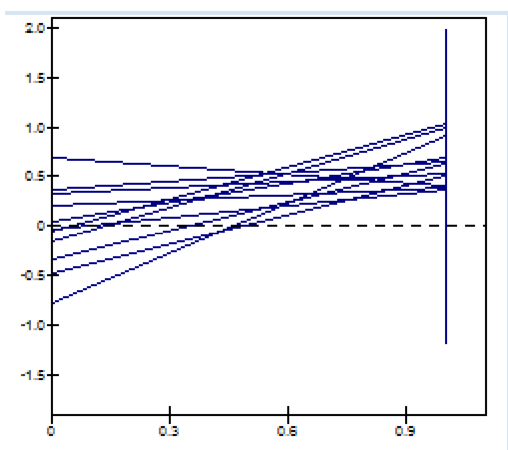
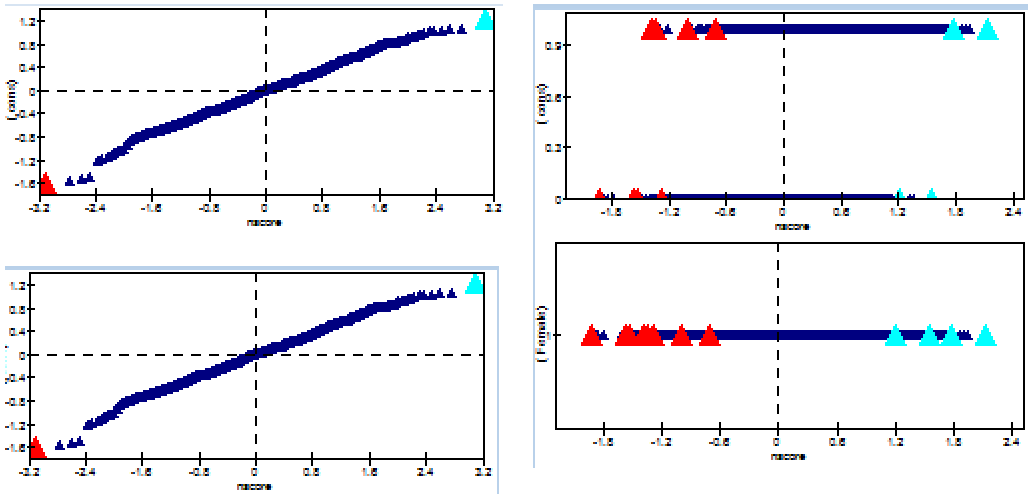



Figure 4.2: Region level (level-3) predicted line fitted by univariate model with random intercept and slope for Sex. From top to bottom the lines corresponding to Tigray, Amhara, Somali, Harari, Beneshangle Gumuzi, SNNP, Gambela, Afar, Addis Ababa, Dire Dawa,and oromiya regions respectively.



(a) (b )

Figure 4.3: (a) Normal probability plot of region level residuals for intercept(top) and Sex (bottom), (b) Region level (level-3) predicted plot for intercept (top) and Sex (bottom).

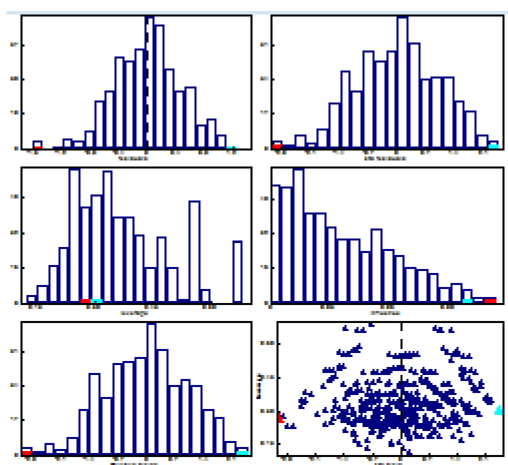


Figure 4.4: Diagnostic plotting: residual, student zed residual, leverage, influence, Deletion residuals for intercept at region level for multilevel multivariate logistic model.