



**ADDIS ABABA UNIVERSITY**

**SCHOOL OF GRADUATE STUDIES**

**Socio-economic and demographic determinants of under-five child  
mortality in Ethiopia**

**Ashagrie Sharew**

**A Thesis Submitted to**

**The Department of Statistics**

**Presented in Partial Fulfillment of the Requirements for the**

**Degree of Master of Science in Statistics**

**Addis Ababa, Ethiopia**

**June, 2013**

Addis Ababa University

School of graduate studies

This is to certify that the thesis prepared by Ashagrie Sharew, entitled: Socio-economic and demographic determinants of under-five child mortality and submitted in partial fulfillment of the requirements for the Degree master of science in Statistics compiles with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Examiner Ebren Legesse signature [Signature] Date 26/06/2013

Examiner M. K. S. Rakoma signature M. K. S. Rakoma Date 26/06/13

Advisor Mekonnen Tadesse signature [Signature] Date 26/06/2013

---

**Chair of Department or Graduate Program coordinator**

## **ABSTRACT**

Socio-economic and demographic determinants of under-five child mortality in Ethiopia

Ashagrie Sharew

Addis Ababa University, 2013

Under-five child mortality level can be used for strategic and national planning as well as a measure of socio-economic indicators like the quality of health care, nutrition, family planning, practice and usage. The story of child survival over the past two decades is one of significant progress and unfinished business. In Ethiopia a significant reduction of under-five child mortality was observed in the last decade though there is much to do to achieve the desire goal. This study aimed to identify the important socio-economic and demographic determinants of under-five child mortality using EDHS 2011 data. Descriptive analysis, cross-tabulation analysis, logistic regression analysis and multilevel regression analysis were used for data analysis. It was found that 7.24 % of under-five children could not survive among children born within five years. Months of breast feeding, mothers' educational level and total children everborn were important factors that affect under-five child mortality. Similarly birth order number of 4-6 and 7-9, number of household members of 6-10 and 11-15 significantly affected under-five mortality. The regions Afar, Amhara, Oromia, Benishangul-Gumuz, SNNP and Gambela were affected most. Children born to mothers of age group 15-19 and 20-24 years were significantly affected based on logistic regression model analysis. The same is true for multilevel regression except that region was a significant random effect. It is recommended that mothers should breast feed their children for at least 6 months; females education enrolment should increase; mothers should give small number of births; mothers should not give birth in their early age ; and regions should be considered while formulating policy to reduce under-five child mortality.

## **Acknowledgement**

First and foremost I would like to thank God for everything. Next I am grateful to Mekonen Tadesse (assistant professor), my thesis advisor, for his tireless support and comments. Without his valuable and encouraging advice, this study would not have been realized.

I finally thank to the Central Statistical Agency for giving me EDHS 2011 data. I also thank my employer Ambo University for supporting me to complete my graduate study successfully.



## **Acronyms**

AIC - Akaike Information Criteria

BIC - Bayesian Information Criteria

Cons-constant

Chi2-chi-square

d.f- degree of freedom

DHS - demographic and health survey

EAs - Enumeration areas

EDHS- Ethiopia demographic and health survey

FDRE- Federal democratic republic of Ethiopia

GDP - Gross Domestic Product

GTP- Growth and Transformation Plan

HIV/AIDS -Human Immune Virus /Acquired Immune Deficiency Syndrome

HSDP - Health sector development programs

ICC- Intra Class Correlation Coefficient

IMR- Infant mortality rate

MDGs - Millennium Development Goals

MOH- Ministry of health

NFHS- National Family health survey

OR - Odds ratio

SAS- Statistical analysis system

S.E- Standard Error

SNNPR- Southern Nations Nationalities and Peoples Region

SNNP-Southern Nations Nationalities and Peoples

SPSS- Statistical Package for Social Science

Sq km- Square kilometer

TV- Television

U5- Under-five

U5M- Under-five mortality

U5MR- Under-five Mortality Rate

UN - United Nations

UNDP -United Nation Development Program

UNICEF- United Nation Children's Fund

USD- United States Dollar

WHO-World Health Organization

## Table of content

|                                                   |     |
|---------------------------------------------------|-----|
| ABSTRACT .....                                    | iii |
| Acknowledgement.....                              | iv  |
| Acronyms .....                                    | v   |
| Table of content.....                             | vii |
| List of tables .....                              | ix  |
| CHAPTER ONE .....                                 | 1   |
| 1. Introduction .....                             | 1   |
| 1.1 Background of the study .....                 | 1   |
| 1.2 Statement of the problem .....                | 9   |
| 1.3 Objective of the study .....                  | 10  |
| 1.3.1 General objective.....                      | 10  |
| 1.3.2 Specific objectives.....                    | 10  |
| 1.4 Significance of the study .....               | 11  |
| 1.5 Limitation of the study .....                 | 11  |
| Chapter two .....                                 | 12  |
| 2. Literature review .....                        | 12  |
| 2.1 Overview of infant and child mortality .....  | 12  |
| 2.2 Situation of infant and child mortality ..... | 13  |
| 2.3 Socio-economic and demographic factors .....  | 15  |
| Chapter three .....                               | 24  |
| 3. Data and methodology .....                     | 24  |
| 3.1 Source of data.....                           | 24  |
| 3.2 Descriptions of variables .....               | 25  |
| 3.2.1 The response variable.....                  | 25  |
| 3.2.2 Explanatory variables.....                  | 26  |
| 3.3 Model description.....                        | 29  |
| 3.3.1 Logistic regression .....                   | 29  |
| 3.3.1.1 Model .....                               | 31  |
| 3.3.1.2 Estimation of model parameters .....      | 32  |
| 3.3.1.4 Residuals for logit models .....          | 36  |

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| 3.3.1.5 Model diagnostic .....                                             | 37 |
| 3.3.2 Overview of multilevel model.....                                    | 38 |
| 3.3.2.1 Multilevel logistic regression .....                               | 41 |
| 3.3.2.1.1 Testing heterogeneous proportions .....                          | 43 |
| 3.3.2.1.2 The empty logistic regression model.....                         | 45 |
| 3.3.2.1.3 The random intercept logistic regression .....                   | 46 |
| 3.3.2.1.4 The Random Coefficient multilevel Logistic Regression Model..... | 49 |
| 3.3.2.1.5 Intra-class Correlation Coefficient (ICC).....                   | 51 |
| 3.3.2.1.6 Parameter estimation .....                                       | 53 |
| 3.3.2.1.7 Significance testing .....                                       | 53 |
| Chapter four .....                                                         | 56 |
| 4. Statistical data analysis and results .....                             | 56 |
| 4.1 Results of descriptive analysis .....                                  | 56 |
| 4.2 Results of cross tabulation data analysis .....                        | 63 |
| 4.3 Logistic regression analysis .....                                     | 64 |
| 4.3.1 Estimation and goodness of fit test.....                             | 64 |
| 4.3.2 Interpretation of parameter estimates .....                          | 68 |
| 4.3.3 Model diagnostics .....                                              | 70 |
| 4.4 Multilevel logistic regression analysis .....                          | 71 |
| 4.4.1 The Empty model with random intercept.....                           | 72 |
| 4.4.2 Random intercept model with fixed coefficients.....                  | 74 |
| 4.4.3 Random intercept with random coefficients model.....                 | 78 |
| CHAPTER FIVE.....                                                          | 83 |
| 5. Discussion, conclusion and recommendation .....                         | 83 |
| 5.1 Discussion and conclusion .....                                        | 83 |
| 5.2 Conclusion.....                                                        | 88 |
| 5.3 Recommendations .....                                                  | 89 |
| References .....                                                           | 90 |
| Appendices.....                                                            | 94 |

**List of tables**

Table3.1. Demographic, socio-economic and proximate variables.....26

Table 4.1 Results of descriptive analysis of socio-economic and demographic factors. ....57

Table 4.2 Results of maximum likelihood estimates of parameters in fitting logistic regression..64

Table 4.3 Result of Model Fit Statistics for Intercept only (empty) and Full Model .....66

Table 4.4: Test of Significance of Hosmer-Lemeshow Goodness of Fit Statistics.....67

Table 4.5 Wald test of all BETAs=0 .....67

Table 4.6 stata xtmelogit output for random intercept only model .....73

Table 4.7 Output of the estimate of random effect for each region.....74

Table 4.8 stata xtmelogit output for random intercept with fixed slope model.....75

Table 4.9 model selection criteria and likelihood ratio test.....76

Table 4.10 stata extmelogit outputs of random slopes model. ....79

Table 4.11 Random-effects correlation matrix for level REGION .....81

Table 4.12 results of model selection criteria and log likelihood ratio test.....82

# **CHAPTER ONE**

## **1. Introduction**

### **1.1 Background of the study**

Ethiopia is the tenth largest country in Africa, covering 1,104,300 square kilometers (with 1.1 million sq km land area and 104,300 sq km water) and is the major constituent of the landmass known as the Horn of Africa. It is bordered on the north-northeast by Eritrea, on the east by Djibouti and Somalia, on the south by Kenya, and on the west-southwest by South Sudan and Sudan. Its geographical coordinates are between 800 N and 3800 E. Ethiopia is a country with great geographical diversity ranging from peaks up to 4,550m above sea level down to depression of 110m below sea level .More than half of the country lies above 1,500 meters. The predominant climate type is tropical monsoon, with three broad climatic variations: the “Kolla”, or hot lowlands, below approximately 1,500 meters, the “Wayna Degas” at 1,500- 2,400 meters and the “Dega” or cool temperate highlands above 2,400 meters. (MEOH, 2010).

Projections from the 2007 population and housing census estimate a total population in 2010 of 79.8 million which makes Ethiopia the second most populous country in Africa next to Nigeria (MEOH, 2010).

According to the 2007 census, it is one of the least urbanized countries in the world with about 83.33% of the population living in rural areas (only 16.67% urban). The largest city in the country is the capital, Addis Ababa, with 2.7 million people (about 4% of the total

population). Nationally, the average household size is 4.7 persons (HSDP IV, 2010/11 – 2014/15)

The population age pyramid has remained predominately young: 44% are under 15 years, over half (52%) are between 15 to 65 years, and only 4% of all persons are over the age of 65 years. The sex ratio between male and female is almost equal, and women in the reproductive ages constitute 24% of the population. While the average lifetime fertility has declined in the past 15 years from a 1990 level of 6.4 births per woman down to 5.4 births (DHS 2005), rural women still have an average of three more births per woman compared to women in urban areas. Overall, even with the fertility decline, the population is still growing at an annual rate of 2.6% where the World has benchmarked a population growth rate of 2% per year as a level beyond which it is difficult for a country's institutions and technologies to keep up with expanding population pressure on all sectors. In Ethiopia, the lowest rate of population growth is in the Amhara region, which, at 1.7%, is lower than population replacement (MEOH, 2010).

Under the 1994 constitution, Ethiopia is a Federal Democratic Republic with nine regional States: Tigray, Afar, Amhara, Benishangul-Gumuz, Gambela, Harari, Oromia, Somali, and Southern Nations Nationalities and Peoples Region (SNNPR); and the City Administration councils of two cities: Dire Dawa and Addis Ababa.

Despite major strides to improve the health of the population in the last one and half decades, Ethiopia's population still face a high rate of morbidity and mortality and the health status remains relatively poor. Vital health indicators from the DHS 2005 show a life expectancy of 54 years (53.4 years for male and 55.4 for female), and an infant

mortality rate (IMR) of 77/1000. Under-five mortality rate has been reduced to 101/1000 in 2010. Although the rates have declined in the past 15 years, these are still very high levels.

But according to United Nations conference on sustainable development (Rio+20) national report of Ethiopia; Infant mortality declined from 97 deaths in 2000 to 59 deaths per 1,000 live births in 2011 (UNICEF, 2012). Similarly, under-five mortality decreased from 166 deaths in 2000 to 88 deaths per 1,000 live births in 2011.

Life expectancy at birth was declining in the 1980s, and dropped sharply from 52 years in 1984 to 46.4 years in 1991, probably because of the HIV/AIDS pandemic, civil war and recurrent famine (Forum for the Environment, 2010.). It has improved over the last two decades, increasing from 46.4 years in 1991 to 59.3 in 2011, according to the UNDP Human Development Report of 2011.

The levels of mortality, especially child or under-five child mortality reflects the state of public health and hygiene, the environmental sanitation, cultural mores about feeding and clothing, socio-economic development, and above all, the people's attitude towards the dignity and value of human life itself. (Bawah, 2001).

Child mortality level can be used for strategic and national planning as well as a measure of socio-economic indicators like the quality of health care, nutrition, family planning, practice and usage. The level provides evidence of policies and practices which lead to the improvement health conditions (Cook *et al.* 2009).

Like GDP, infant and child mortality are foremost indicators of population wellbeing.



Story of child survival over the past two decades is one of significant progress and unfinished business.

More children now survive their fifth birthday than ever before. The global number of under-five deaths has fallen from around 12 million in 1990 to an estimated 6.9 million in 2011. All regions have shown steady reductions in under-five mortality over the past two decades. In the last decade alone, progress on reducing child deaths has accelerated, with the annual rate of decline in the global under-five mortality rate rising from 1.8% in 1990-2000 to 3.2% in 2000-2011 (UNICEF, 2012).

The gains have been broad, with marked falls in diverse countries. Between 1990 and 2011, nine low-income countries: Bangladesh, Cambodia, Ethiopia, Liberia, Madagascar, Malawi, Nepal, Niger and Rwanda - reduced their under-five mortality rate by 60% or more (UNICEF, 2012).

The advances to date stem directly from the collective commitment, energy and efforts of governments, donors, non-governmental organizations, UN agencies, scientists, practitioners, communities, families and individuals. Measles deaths have plummeted. Polio, though stubbornly resistant thus far to elimination, has fallen to historically low levels. Routine immunization has increased almost everywhere. Among the most striking advances has been the progress in combating AIDS. Thanks to the application of new treatments, better prevention and sustained funding, rates of new HIV infections, and HIV-associated deaths among children have fallen substantially (UNICEF, 2012).

But any satisfaction at these gains is tempered by the unfinished business that remains. The fact remains that, on average, around 19,000 children still die every day from largely preventable causes. With necessary vaccines, adequate nutrition and basic medical and maternal care, most of these young lives could be saved (UNICEF, 2012).

Nor can we evade the great divides and disparities that persist among regions and within countries. The economically poorest regions, least developed countries, most fragile nations, and most disadvantaged and marginalized populations continue to bear the heaviest burden of child deaths. More than four-fifths of all under-five deaths in 2011 occurred in sub-Saharan Africa and South Asia (UNICEF, 2012).

In June 2012, the Governments of Ethiopia, India and the United States together with UNICEF brought together more than 700 partners from the public, private and civil society sectors for the Child Survival Call to Action. What emerged from the Call to Action was a rejuvenated global movement for child survival, with partners pledging to work together across technical sectors with greater focus, energy and determination. Since June, more than 110 governments have signed a pledge vowing to redouble efforts to accelerate declines in child mortality; 174 civil society organizations, 91 faith-based organizations, and 290 faith leaders from 52 countries have signed their own pledges of support (UNICEF, 2012).

Reductions in under-five mortality rates, combined with declining fertility rates in many regions and countries, have diminished the burden (number) of under-five deaths from nearly 12 million in 1990 to an estimated 6.9 million in 2011. About 14,000 fewer children die each day than did two decades ago. A testimony to the sustained efforts and

commitment to child survival by many, including governments and donors, non-governmental organizations and agencies, the private sector, communities, families and individuals (UNICEF, 2012).

Mortality rates among children under 5 years of age fell globally by 41% between 1990, the base year for the Millennium Development Goals (MDGs), and 2011, lowering the global rate from 87 deaths per 1,000 live births to 51. Importantly, the bulk of the progress in the past two decades has taken place since the MDGs were set in the year 2000, with the global rate of decline in under-five mortality accelerating to 3.2% annually in 2000-2011, compared with 1.8% for the 1990-2000 periods (UNICEF, 2012).

Even as the global and regional rates of under-five mortality have fallen, the burden of child deaths has become alarmingly concentrated in the world's poorest regions and countries. A look at how the burden of under-five deaths is distributed among regions reveals an increasing concentration of mortality in sub-Saharan Africa and South Asia; in 2011, more than four-fifths of all global under-five deaths occurred in these two regions alone. Sub-Saharan Africa accounted for almost half (49%) of the global total in 2011. Despite rapid gains in reducing under-five mortality, South Asia's share of global under-five deaths remains second highest, at 33% in 2011. In contrast, the rest of the world's regions have seen their share fall from 32% in 1990 to 18% two decades later.

The highest regional rate of under-five mortality is found in sub-Saharan Africa, where, on average, 1 in 9 children die before age 5. (UNICEF, 2012)

In 2000, the Ethiopian Demographic and Health Survey (EDHS) estimated the under five mortality (U5MR) at 166 per 1000 live births. The Ethiopian Ministry of health (MOH) estimated that the U5MR for 2002-2003 was 140. In 2000, the rate of 166 placed Ethiopia 21<sup>st</sup> in the world for under-five mortality.

From 1960 to 2000 there has been an average 1.2% annual decrease in under-five mortality.

According to the child health in Ethiopia (2004) report, among under- five deaths, 29% die in the first 30 days of life, 29% from the first to the 11<sup>th</sup> month and 42% from 1 to 4 years of age.

Based on UNICEF, 2012 report the U5MR declined from 198/1000 in 1990 to 139/1000 in 2000 and to 77/1000 with the millennium development goal of 66 in 2015.

It has been many years since world leaders committed to Millennium Development Goal 4 (MDG 4), which sets out to reduce the under-five mortality rate by two-thirds between 1990 and 2015. The world has made substantial progress, reducing the under-five mortality rate 41 percent, from 87 deaths per 1,000 live births in 1990 to 51 in 2011. However, this progress has not been enough, and the target risks being missed at the global level. The global under-five mortality rate needs to be reduced to 29 deaths per 1,000 live births which implies an annual rate of reduction of 14.2 percent for 2011–2015, much higher than the 2.5 percent achieved over 1990–2011. Still, in 2011, 6.9 million children died before reaching their fifth birthday. Almost two-thirds of them, 4.4 million, died of infectious diseases, nearly all of which were preventable. In addition, inequities in child mortality between developing and developed regions remain large.

The government of Ethiopia has been implementing four health sector development programs (HSDP) since 1997/1998 each of which incorporates reduction of infant and child mortality rates (MEOH, 2010).

Through the progressive implementation of these HSDPs, Ethiopia has made great strides to improve infant and child survival.

In 2000 the Ethiopian government announced the intention by signing the millennium declaration committing to achieve the Millennium Development Goals (MDGs) by 2015, many of which overlap with the 2015 national policy goals, which introduced by the federal government in 1991 and a policy action has been continued, for instance, in 2004 the Ethiopian government prepared child survival strategy and implementation plan to reduce under-five mortality of 140/1000 live births to 67/1000 live births by 2015, this means a reduction of two-thirds from the 1990 rates of about 200/1000 live births or a 52 percent reduction from the 2004 rate of about 140/1000 live births (Federal Ministry of health, 2005).

The Government of Ethiopia has embarked on a new five year plan (2011-2015), called the Growth Transformation Plan (GTP). The GTP aims to foster broad-based development in a sustainable manner to achieve the Millennium Development Goals. The Plan envisages a major transformation of the economic structure, seeking to double agricultural production and significantly increase the share of industry in the economy. The plan seeks to reduce infant mortality rates from 101 per 1000 to 67 per 1000 and cut the maternal mortality rate by more than half from 590 per 100,000 to 267 per 100,000. (United Nations Country Team, March 2011)

The study of infant and child mortality becomes one of the most important researches of the developing countries including Ethiopia. Child mortality is a composite index reflecting environmental, social, economic, health care services and delivery situation on the one hand and maternal as well as family and community norms and practices on the other.

Numerous studies have shown a close association between child mortality and socio-economic status (for example, Preston, 1975 and 1985; Hobcraft et al, 1984; Hill, 1985; World Bank, 1993). Most indicators of socio-economic status used are income per capita, education, urban/rural residence, work status and household assets.

## **1.2 Statement of the problem**

Under-five child mortality decreased significantly since 1960 but there is still unfinished works to be done. That is why the world leaders set the millennium development goals (MDGs) which are being implemented from 1990 to 2015. One of these goals is reduction of infant and child mortality by two-third by 2015.

Under-five child mortality is high all over the world on average. Particularly the problem is worse in sub-Saharan countries that accounted for almost half (49%) of the global total in 2011.

Ethiopia has high infant and child mortality though pronounced reduction is recorded for the last three decades. This reduction is accounted by the implementation of successive health sector development programs and increasing allocation of budget although still well below the World Health Organization's (WHO) standard.

Nowadays under-five child mortality is a great concern for both the governments and researchers. Many researchers are being attracted to identify the major factors associated with infant and child mortality and are giving their own recommendations.

Researchers agreed that pneumonia, malaria, diarrhea, measles, neonatal causes (prematurity, asphyxia and neonatal sepsis), malnutrition, and HIV/AIDS are underlying causes of infant and child mortality. But these causes are worsened by socio-economic and demographic determinants.

Though Ethiopia is one of the countries experiencing high under-five child mortality, researches done concerning factors affecting under-five child mortality are not enough. Especially socio-economic and demographic factors which are affecting under-five child mortality more significantly are not well studied.

Consequently, this study attempts to identify some socio-economic and demographic factors affecting under-five child mortality. It also examines whether there are variations in under-five mortality among regions of Ethiopia.

### **1.3 Objective of the study**

#### **1.3.1 General objective**

The general objective of the study was to identify the most significant socio-economic and demographic factors affecting under-five child mortality and evaluate the variation of these factors within and between regions.

#### **1.3.2 Specific objectives**

- To identify socio-economic and demographic factors that affect most under-five child mortality.

- To determine the extent of under-five child mortality within and between regions.
- To identify the factors that may explain the variation in infant and child mortality between regions of Ethiopia.

#### **1.4 Significance of the study**

This study aimed to identify the levels and trends of under-five child mortality and associated socio-economic and demographic factors influencing under-five mortality.

Hence the finding of this study could be useful in many respects.

The results of the study could also be helpful in pointing out the regional variation of under-five child mortality and suggesting the solution to minimize this disparity. Policy maker, monitors and evaluators could benefit to evaluate their policies and programs towards reducing under-five mortality.

Moreover, researchers, governmental and non-governmental organizations who are working on infant and child mortality could use the findings..

#### **1.5 Limitation of the study**

1. This study used a retrospective data where misreporting of events could occur due to lack of memory of the respondent.
2. The survey took 11,654 respondents but due to missing values, inconsistent answers and blank values of each variable, the study used 11,250 respondents.
3. Birth interval could be an important variable for under-five child mortality but because of large number of missing values this indicator was not included in the study.



## **Chapter two**

### **2. Literature review**

#### **2.1 Overview of infant and child mortality**

Under-five mortality has declined since 1960s all over the World even though there are regional disparities. But it still remains to be one of World's major issues. That is because World leaders set out millennium development goals and the target for one of these millennium development goals (MDG4) is reduction of infant and child mortality rates by 67% between 1990 and 2015 (UNICEF,2012).

Now substantial progress has been made to achieve MDG4 but it is challenging for most developing countries in particular Sub-Saharan Africa where Ethiopia is found.

The highest mortality rates are in Sub-Saharan Africa where 1 in 9 children die before age five, more than 16 times the average for developed regions (1 in 152) and southern Asia (1 in 16). Although U5MRs have fallen more sharply elsewhere, the disparity between these two regions and the rest of the World has grown (UNICEF; 2012).

This disparity can be caused by lack of commitment by the government and the socio-economic and demographic inequality which shows the variation in education, level of income, lack of infrastructure, lack of awareness for family planning and so on (World Health Statistics, 2012).

According to UNICEF, analysis of international household survey data shows that children born into the poorest quintile (fifth) of households are almost twice as likely to die before age 5 as their counterparts in the wealthiest quintile. Poverty is not the only

divider, however. Children are also at greater risk of dying before age 5 if they are born in rural areas, among the poor, or to a mother denied basic education (progress report 2012 by UNICEF'S division of policy and strategy; renewing the promise-in every country, for every child)

## **2.2 Situation of infant and child mortality**

Between 1990 and 2006, the number of child deaths globally was cut from 13 to just fewer than 10 million per year. The vast majority of these deaths occurred in developing countries and from preventable causes. Globally, significant progress has been made in reducing mortality rates among children under five years old. Between 1990 and 2010, the under-five mortality rate declined by 35% from an estimated 88 deaths per 1000 live births to 57. The global rate of decline has also accelerated in recent years from 2.1% per annum during 1990–2010 to 2.6% during 2005–2010 (World health statistics, 2012).

According to UNICEF (2012) substantial progress has been made towards achieving MDG 4. The number of under-five deaths worldwide has declined from nearly 12 million in 1990, the baseline for MDG, to 6.9 million in 2011. While that translates into 14,000 fewer children dying every day in 2011 than in 1990, it still implies the deaths of 19,000 children under age five every day in 2011.

The annual rate of reduction in under-five mortality has accelerated from 1.8 percent a year over 1990–2000 to 3.2 percent over 2000–2011 but remains insufficient to reach MDG 4, particularly in Oceania, Sub-Saharan Africa, Caucasus and Central Asia, and Southern Asia.

The highest rates of child mortality are still in Sub-Saharan Africa where 1 in 9 children dies before age five, more than 16 times the average for developed regions (1 in 152) and Southern Asia (1 in 16). As under-five mortality rates have fallen more sharply elsewhere, the disparity between these two regions and the rest of the world has grown (UNICEF, 2012).

Under-five deaths are increasingly concentrated in Sub-Saharan Africa and Southern Asia, while the share in the rest of the world dropped from 31 percent in 1990 to 17 percent in 2011. The share of under-five deaths that occur in Sub-Saharan Africa and Southern Asia is large and growing (83% in 2011). Of the 24 countries with under-five mortality rate above 100 deaths per 1,000 live births in 2011, 23 are in Sub-Saharan Africa and the remaining one is in Southern Asia. As under-five mortality rates have fallen more sharply in richer developing regions, the disparity between Sub-Saharan Africa and other regions has grown. In 1990 a child born in Sub-Saharan Africa faced a probability of dying before age five that was 1.5 times higher than in Southern Asia, 3.4 times higher than in Latin America and the Caribbean, 3.7 times higher than in Eastern Asia and 12.1 times higher than in developed regions. By 2011 that probability was 1.8 times higher than in Southern Asia, 5.7 times higher than in Latin America and the Caribbean, 7.4 times higher than in Eastern Asia and 16.5 times higher than in developed regions. However, Sub-Saharan Africa has seen a faster decline in its under-five mortality rate with annual rate of reduction doubling between 1990-2000 and 2000-2011 (UNICEF, 2012)

Ethiopia being one of Sub-Saharan African, that experience high infant and child mortality, is still in trouble with under-five mortality. Based on the mortality trend obtained from DHS survey conducted in 2000, 2005 and 2011, it is evidenced that infant and child mortality in Ethiopia declined continuously. Under-five mortality decreased from 166 deaths per 1,000 live births in 2000 to 59 deaths per 1,000 live births in 2011 (EDHS; 2011).

### **2.3 Socio-economic and demographic factors**

Socioeconomic conditions of people in the community have strong impact on infant and child mortality. The relationship between socioeconomic development and mortality level is circular in the sense that socioeconomic development affects the mortality level and mortality level affects socioeconomic and demographic structure of a country. In the developing countries, several socioeconomic factors have been identified as having strong relationship with infant and child mortality differentials. The type of place of residence is also an important determinant of child survival. Kabir and Chowdhury (1993) explained that urban-rural differentials may be attributed to different health care services including higher coverage with immunization, safe delivery of births and access to health care services.

Mondaal (2009) using data from Strengthening the Department of Population Science and Human Resource Development with cross classification and logistic regression have shown that place of residence, mother's education, father's education, mother's age at birth, birth interval, breast feeding, having television and access to facility have significant impact on infant and child mortality. The father's educational attainment affects infant and child mortality because of his decision giving power in most

developing countries. Similarly father's occupation is an important predictor of infant and child mortality. One's occupation is reflection of the physical environment, his social milieu, his education background, his income and his life style.

Mother's occupation is also associated with nutritional status of their children. The incidence of infant and child mortality is higher among not working women including regular house hold activities than that of working women. Television (TV) is at present a powerful media and basic thing of life. When an illiterate man watches health related programs, he will realize the importance and role of health and cleanness.

On the other hand Arvind Pandey (1998) using data from India's National Family Health Survey (NFHS) showed that urban/rural residence, mother's exposure to mass media, mother's literacy, access to flush or pit toilet, house hold's religion, mother's age at the time of birth and economic level of the household have substantial effect on infant and child mortality and are statistically significant. Among children born to mothers at various ages, those born to mothers under the age of 20 have the least mortality rates. They also showed that for children who are not first born, previous birth interval has by far the largest effect on infant and child mortality. Children born less than 24 months after a previous birth are more than twice as likely to die during infancy and two thirds more likely to die during childhood compared with children born after a longer interval. Similarly they said that families that have already experienced the death of an infant or child are at much greater risk than other families.

Literate mothers have more than double advantage as compared to their illiterate or uneducated counter parts of not being sufferer of their babies' early death. Other factors being constant, maternal education is one of the leading determinants of infant and child

mortality and is a principal source of ascertaining a health seeking behavior of mother towards their children (Safeer Ullah, 2010). This is the same as Caldwell's (1979) who demonstrated that mother's education is a more decisive determinant of child survival than other family characteristics such as husband's occupation and his education.

Mother's education affects child survival in two ways: through better child-care practices and higher standards of hygiene at home and more rational and greater use of preventive and curative medical services (Mosely and Chen, 1984; Cleland and Van Ginneken 1988).

Kwabena (2011) based on data from South Africa demographic and health survey and using logistic regression analysis has found that education of the mother, marital status of the mother, the number of children the mother has, duration of breast feeding, and previous loss of a child are determinants of under-five mortality. But in contrary to the aforementioned results he found that father's occupation and availability of toilet have no significant impact for under-five survival.

Tawiah (1979) has shown that work status of a mother is an important determinant of mortality of her children at the younger ages, zero to five. The study explains that when a mother works outside the home, she is prevented from taking maximum care of the child and that affects the survival chances of the child directly. On the other hand, the longer the duration of breastfeeding (from HIV negative women) the lower the risk of child mortality; the risk reduced by over 33 percent. The study showed that women's education after nine successful years of schooling is a necessary and essential factor to promote a decline in child mortality.

Though not expected, the influence of mother's education on infant mortality completely disappears from 2005-2006 Zimbabwe DHS (Joshua Kombo and Jeroen K. Van Ginneken, 2009). Using multivariate analysis they found no-U shaped relationship between birth order and mortality and maternal age and mortality. In the mean time they found that lower risks of child death among children who are first born and those born to mothers aged 40-49 years which is a deviation from what others agreed. But they revealed that birth order, preceding birth interval and maternal age as determinants of infant mortality, which are less pronounced in child mortality. The provisions of piped drinking water and flush toilets to households have a strong impact on child mortality than infant mortality.

Trussel and Hammerslough (1983) in hazard model analysis for child mortality in Sri Lanka, found that improvement in the type of toilet facility reduces mortality while the source of water supply was found insignificant. Ridder and Tunali (1999) found that access to piped water and toilet facilities did not have much of an impact on the child mortality risk in Malaysia.

Mturi and Curtis (1995), based on data from 1991/92 DHS in Tanzania, used hazard model and found that short birth interval, adolescent pregnancy, previous child mortality are associated with increased risk of infant and child mortality while no significant effect of socio-economic status (maternal education, partner's education, urban/rural residence and presence of radio in the house hold) of the population on infant and child mortality was observed.

Similarly, Mustafa and Odimegwu(2008), based on Kenya's DHS data, used logistic regression and found similar result as that obtained in Tanzania for socio-economic

determinants of infant and child mortality rate both in urban and rural areas. They found that regional variation exist in infant and child mortality between the different provinces of Kenya.

According to World bank's country status report (June, 2004), infant and under-five mortality rates in Ethiopia are 16 percent and 31 percent higher respectively among children from the poorest quintile compared to children from richest quintile. In urban and rural areas infant mortality rates (IMR) are 96.5 and 114.7 and under -five mortality rates are 148.6 and 192.5 respectively which shows there is a remarkable difference between urban and rural residence. Besides mortality rates vary between regions: Addis Ababa has the lowest IMR (81) and U5MR (113) while Gambella has the highest IMR (123) and U5MR (223).

In this report it was stated that short birth intervals, a high birth order, a low birth weight, the young age of mothers, and being part of certain religious groups are all strongly correlated with high child mortality levels. Low child mortality rate was among Orthodox Christians and low mortality rate among children of older women.

The government of Ethiopia, using EDHS (2005) data has stated the following in the report on progress in implementing the World Fit for children plan of action in Ethiopia (2007). There are significant regional variations in infant and U5 mortality that reflect regional disparities, urban-rural differences and educational and wealth levels. For instance U5MR ranges from a low of 72 in Addis Ababa to a high of 157 per 1,000 live births in Benishangul Gumuz followed by Gambella (156) and Amhara Regions (154). U5MR among children born to mothers with no education (139 per 1,000 live births) is more than twice that of children born to mothers with secondary and higher level of



education (54 per 1,000 live births). The U5MR for children who live in rural areas is 135 per 1000 live births versus 98 per 1000 live births for those who live in urban areas.

According to EDHS (2011) report infant and child mortality are strongly influenced by socio-economic and demographic differentials. Mortality rates are consistently lower in urban than in rural areas. Infant mortality is 29% higher in rural areas (76 deaths per 1000 live births) than in urban areas (59 deaths per 1,000 live births). The urban-rural difference is also more pronounced in the case of U5M. U5MRs range from a low of 53 per 1,000 live births in Addis Ababa to a high of 169 per 1000 live births in Benishangul-Gumuz. U5M is also high in Afar, Gambella and Somali. Mother's education is inversely related to a child's risk of dying. U5M among children born to mothers with no education (121 per 1,000 live births) is 2.6 times as high as that of children born to mothers with secondary education (46 per 1,000 live births) and more than five times as high as that of mothers having above secondary education (24 per 1,000 live births). The beneficial effect of educating mothers is evident for all childhood mortality rates. Also, childhood mortality generally decreases as wealth increases, although the differences are relatively small in the three highest wealth quintiles.

Sex of the child, mother's age at birth, birth order, previous birth interval and the infant size at birth have been found to play an important role in the survival of children.

Kumar and Gemechis (2010) used cross tabulation technique on EDHS (2005) data and found that birth interval with previous birth and mother's education have a significant impact on child mortality. The result confirmed that the child mortality risk of children of less than 2 years of birth interval with previous child was highest (15%) and lowest

(4.2%) for the children whose birth interval was 4+ years. On the other hand they reported that children's mother's educational level is significantly correlated with the risk of mortality of children.

Desta Mekonnen (2011) used data from EDHS (2005) to investigate the role of Socioeconomic, Demographic and Biological factors in the previous five years period of 2000 and 2005 by employing logistic regression method. He found that marital status, birth order, type of birth and preceding birth interval are the dominant significant proximate determinants of infant and child mortality. Breast feeding is also found to be significant for infant mortality but it is insignificant for child mortality. Children born to women not currently married, first born children, children born after less than 18 months of previous birth and children who breast fed less than 6 months are exposed to high risk of mortality. On the other hand it was found that among socio-economic variables, household size, and to some extent mother's and father's educational level and sex of household head are the most significant determinants of infant and child mortality. Children born in households of small size, children born in households headed by males, children born to mothers and fathers with no education and to some extent children born to mothers with primary education are exposed to high risk of infant and child mortality.

## **2.4 Research hypotheses**

Death is a natural process which is caused by illness but the cause of illness in under-five is aggravated with socio-economic, demographic and other proximate determinants enhancing the death of infants and children. Even though substantial reduction of infant and child mortality is recorded in Ethiopia for the last few decades as mentioned above,

much is left to achieve the desired goal especially the MDG4 which is the promise of reduction of U5M by two-third by 2015.

Thus, based on previous research findings and theories the following hypothesis will be explored and factors more associated with infant and child mortality in Ethiopia will be identified.

1. The risk of infant and child mortality is expected to be lower among children whose mothers have at least secondary level education.
2. The risk of infant and child mortality is expected to be lower for women having less number of children.
3. The risk of infant and child mortality will be higher in rural areas than in urban areas.
4. The risk of infant and child mortality is expected to be higher for uneducated father than the educated father.
5. The risk of infant and child mortality is expected to be higher for young and old mothers.
6. Infant and child mortality is expected to be lower if there is television (TV) in the household.
7. Infant and child mortality is expected to be different among different religions.
8. Infant and child mortality is expected to vary between and across regions due to different socio-economic activities, cultures and uneven distribution of infrastructures.
9. The risk of under-five child mortality is expected to decrease as months of breast feeding increases.

10. Under-five child mortality is expected to be higher for mothers whose age at first birth is less than 20.
11. The under-five child mortality is assumed to be increasing as the number of children born to the mother increases.

These aforementioned and other unstated hypotheses will be verified by using ordinary logistic regression and multilevel logistic regression models.

## **Chapter three**

### **3. Data and methodology**

#### **3.1 Source of data**

The data of this study were obtained from Ethiopian DHS conducted in 2011. DHS is a worldwide project and EDHS was conducted by Ethiopian central statistical agency (CSA) under supervision of the government of Ethiopia funded by United States agency for international development.

The 2011 EDHS sample was selected using a stratified, two-stage cluster design, and enumeration areas (EAs) were the sampling units for the first stage and each household was the sampling unit of the second stage. The sample included 624 EAs, 187 in urban areas and 437 in rural areas. A total of 17,817 households were selected for the sample, of which 17,018 were found to be occupied during data collection. Of these, 16,702 were successfully interviewed, yielding a household response rate of 98 percent. In the interviewed households 17,385 eligible women aged 15-49 were identified for individual interview; complete interviews were conducted for 16,515, yielding a response rate of 95 percent (CSA, 2011).

The data used for mortality estimation were collected in the birth history section of the Woman's Questionnaire. The birth history section begins with questions about the respondent's experience with childbearing (i.e., the number of sons and daughters living with the mother, the number who live elsewhere, and the number who have died). These

questions are followed by a retrospective birth history, in which each respondent is asked to list each of her births, starting with the first birth. For each birth, data were obtained on sex, month and year of birth, survivorship status, and current age, or, if the child is dead, age at death. This information is used to directly estimate mortality rates.

The EDHS 2011 data set has hierarchical structure. The hierarchy for this study follows individuals/women as level-1, and regions as level-2. This means that individuals are nested in regions.

### **3.2 Descriptions of variables**

The dependent and independent variables are selected based on previous works to analyze and see which independent or explanatory variables affected the dependent or response variable significantly by using both descriptive and inferential analysis.

#### **3.2.1 The response variable**

The response variable of the study is under-five child mortality (coded 0 or 1) in Ethiopia for the period of five years from 2005-2011. For the purpose of clarity, it is important to define different stages of mortality according to time. Infant mortality is the probability of dying in the age group 0-1 year. Child mortality is the probability of dying in the age group 1-5 years. Under-five child mortality is the probability of dying in the age group 0-5 years where mortality is measured per 1000 live births.

The response variable is dichotomous and is denoted by  $Y_i$ . If a mother experienced a loss of under-five child in the last five years before the interview, the response variable ( $Y_i$ ) takes the value of 1 otherwise it takes the value of 0.

In short:  $Y_i = \begin{cases} 1 & \text{if the } i\text{th mother had the incidence of U5CM} \\ 0 & \text{if the } i\text{th mother had not experienced U5CM} \end{cases}$

### 3.2.2 Explanatory variables

The explanatory variables explored in this study were based on previous works done especially according to Mosley and Chen (1984) framework. As determinants of under-five child mortality, this study utilized only socio-economic and demographic factors. These explanatory variables are educational level of mother, occupation of mother, the number of children ever born to the mother, marital status of mother, place of residence, occupation of husband/partner, education of husband/partner, months of breast feeding, owning TV or not, age of mother at first birth, sex of child, religion, region, birth order number, wealth index, number of household members, age of mother and sex of household head.

Table3.1. Demographic, socio-economic and proximate variables.

| Variables | Descriptions | Values/records                                                                                                                                     |
|-----------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| 1.REGION  | Region       | 1=Tigray<br>2=Afar<br>3=Amhara<br>4=Oromia<br>5=Somali<br>6=Benshangule-Gumuz<br>7=SNNP<br>8=Gambella<br>9=Harari<br>10=Addis Ababa<br>11=Diredawa |
| 2.RELGION | Religion     | 1=Orthodox<br>2=Catholic<br>3=Protestant<br>4=Muslim<br>5=Traditional                                                                              |

|            |                                        |                                                                                                                            |
|------------|----------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
|            |                                        | 6=Other                                                                                                                    |
| 3.PR       | Type of Place of residence             | 1=urban<br>2=rural                                                                                                         |
| 4.MEDU     | Mother's educational level             | 0=no education<br>1=primary education<br>2=secondary<br>3=Higher                                                           |
| 5.MOCCUP   | Mother's occupation                    | 0=not working<br>1=Professional/technical/managerial/clerical<br>2=Sales/services<br>3=Agricultural<br>4= Manual           |
| 6.HEDU     | Husband/partner's education            | 0=no education<br>1=primary education<br>2=secondary<br>3 =higher<br>8= Don't know                                         |
| 7.HOCCUP   | Husband/partner's occupation           | 0=Jobless<br>1=Managers<br>2=Agricultural worker<br>3=Trade worker<br>4=Operator/assembler<br>5=Labourer<br>6=Other        |
| 8.TCHBORN  | Total children ever born to the mother | 1=1-4<br>2= 5-8<br>3=9 and above                                                                                           |
| 9.CMSTATUS | Current marital status                 | 0=Never in union<br>1=Married<br>2=Living with partner<br>3=Widowed<br>4=Divorced<br>5=No longer living together/separated |
| 10.NHM     | Number of household members            | 1=1-5<br>2=6-10<br>3=11-15<br>4=16 and above                                                                               |



|                 |                              |                                                                                                                |
|-----------------|------------------------------|----------------------------------------------------------------------------------------------------------------|
| 11.MBFEED       | Months of breast feeding     | 1=0-6<br>2=7-12<br>3=13-18<br>4=19-59<br>5=ever breast fed, not currently breast feeding<br>6=never breast fed |
| 12. HTV         | Household has Television     | 0=no<br>1=yes                                                                                                  |
| 13. SEXH        | Sex of household head        | 1=male<br>2=female                                                                                             |
| 14.<br>AGEFIRST | Age of mother at first birth | 1=9-14<br>2=15-19<br>3=20-24<br>4=25-29<br>5=30 and above                                                      |
| 15.SEX          | Sex of child                 | 1=male<br>2=female                                                                                             |
| 16.BORD         | Birth order number           | 1=1-3<br>2=4-6<br>3=7-9<br>4=10 and above                                                                      |
| 17.AGEM         | Age of mother                | 1=15-19<br>2=20-24<br>3=25-29<br>4=30-34<br>5=35-39<br>6=40-44<br>7=45-49                                      |
| 18.WINDEX       | Wealth index                 | 1=poorest<br>2=poorer<br>3=middle<br>4=richer<br>5=richest                                                     |

Note: In the category of months of breastfeeding, 0-6,7-12,13-18 and 19-59 months are the categories for currently breast feeding.

### **3.3 Model description**

#### **3.3.1 Logistic regression**

Regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. The outcome variable in binary logistic regression is binary or dichotomous while the outcome variable in linear regression is continuous. On the other hand the error term in linear regression is normally distributed with mean zero and some variance that is constant across levels of the independent variables while the error term in binary logistic regression is distributed binomially, not normally. Consequently the response variable in logistic regression is not restricted to normality in case of parameter estimation. Because of this logistic regression is the most popular method of analyzing binary response data.

There are two primary reasons for choosing the logistic regression. First from mathematical point of view it is an extremely flexible and easily used function and second it lends itself to a clinically meaning full interpretation (Hosmer and Lemeshow, 2000).

Logistic model, as compared to its competitor, the probit model, is less sensitive to outliers and easy to correct a bias (Copas, 1988). Logistic regression is used regularly rather than discriminate analysis when there are two categories of the dependent variable. Discriminate analysis strictly requires the continuous independent variables (though dummy variables can be used as in multiple regressions). Thus, in instances where the independent variables are categorical, or a mix of continuous and categorical, and the dependent variable is categorical, logistic regression is necessary (Agresti, 2007).

Logistic regression determines the impact of multiple independent variables presented simultaneously to predict membership of one or other of the two dependent variable categories.

There are two main uses of logistic regression: the first is the prediction of group membership. Since logistic regression calculates the probability of success over the probability of failure, the results of the analysis are in the form of an odds ratio. Logistic regression also provides knowledge of the relationships and strengths among the variables.

The assumptions required for statistical tests in logistic regression are far less restrictive than those for ordinary least squares regression. There is no formal requirement for multivariate normality, homoscedasticity, or linearity of the independent variables within each category of the response variable. However, the assumptions that apply to logistic regression model include: meaningful coding, inclusion of all relevant and exclusion of all irrelevant variables in the regression model and low error in the explanatory variables.

**Odds ratio:** for a probability of success  $\pi$ , the odds of success are defined to be the ratio of probability of success  $\pi$  (in our case the probability of under-five child mortality or  $y=1$ ) to probability of failure  $1 - \pi$  (in our case the probability of survival of under-five or  $y=0$ ).

$$\text{Odds} = \frac{\pi}{1 - \pi} \quad (3.1)$$

The odds are non-negative with value greater than one when a success is more likely than a failure and between 0 and 1 when the appropriate is true. The odds ratio denoted as

OR is defined as the ratio of the odds for  $y=1$  to the odds for  $y=0$  and is given by the equation:

$$OR = \frac{\pi(1)}{1-\pi(1)} * \frac{1-\pi(0)}{\pi(0)} \quad (3.2)$$

The odds ratio is a measure of association which has found wide use, as it approximates how much more likely (or unlikely) it is for the outcome to be present among those with  $y=1$  than among those with  $y=0$ .

### 3.3.1.1 Model

Let  $Y_{n \times 1}$  be a dichotomous outcome random vector with categories 1 (a woman experienced under-five mortality) and 0 (a woman didn't experience under-five mortality). Let  $X$ , an  $n \times (k+1)$  matrix denotes the collection of  $k$ -predictor variables and,  $\beta$  be a  $(k+1) \times 1$  vector of parameters, then the data layout of explanatory variables is given by:  $X = (1, x_1, x_2, \dots, x_k)'$   $\beta = [\beta_0, \beta_1, \dots, \beta_k]'$

The first (leading) column is a column of 1's. It corresponds to the constant or intercept of logistic regression equation.  $X$  without the leading column of 1s, is termed as predictor data matrix. Then, the conditional probability of  $Y$  given  $X$  is given by:

$$\Pi = p(Y = 1|X) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} = \frac{\exp(X'\beta)}{1 + \exp(X'\beta)} \quad (3.3)$$

From (3.3) the odds of success is given by:

$$Odds(Y = 1) = \frac{\pi}{1-\pi} = \exp(X'\beta) \quad (3.4)$$

In logistic regression analysis, it is assumed that the explanatory variables affect the response variable through a suitable transformation of odds of the probability of

success. This transformation is a suitable link function of  $\pi$  and is called the logit - link, which is defined as:

$$\text{Logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \log(\exp((\mathbf{X}'\boldsymbol{\beta})))$$

The transformed variable, denoted by  $\text{logit}(\pi)$  is the log-odds and is related to the explanatory variables as:

$$\text{Logit}(\pi) = g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \mathbf{X}'\boldsymbol{\beta} \quad (3.5)$$

The above equations give suitable representations of the success probability, odds, and log-odds. Indeed, these representations facilitate interpretations of parameter estimates.

The parameter  $\beta_i$  refers to the effect of  $X_i$  on the log- odds for  $Y = 1$  controlling the other  $X$ 's in the model. For predictor variables having  $L$  levels ( $L \geq 2$ ), interpretation can be made by making one of the  $L$ -levels as a reference category.

### 3.3.1.2 Estimation of model parameters

The logistic regression model just developed is a generalized linear model with binomial errors and logit link. Instead of finding the best fitting line by minimizing the sum of the squared residuals, as we did with OLS regression, we use Maximum Likelihood (ML) estimation to obtain estimates the coefficients in the logit equation given in (3.5).

ML methods seek to maximize the log likelihood,  $LL$ , which reflects how likely it is (the odds) that the observed values of the outcome may be predicted from the observed values of the predictors. ML method for this case is based on Newton-

Raphson iteratively reweighted least square algorithm until it gets the smallest possible deviance or best fit.

Suppose  $(y_1, y_2, \dots, y_n)$  represent the  $n$  dependent random observations corresponding to the random variables  $(Y_1, Y_2, \dots, Y_n)$ . Since the  $Y_i$  is a Bernoulli random variable, the probability function of  $Y_i$  is  $f(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$ ;  $y_i = 0$  or  $1$   $i=1, 2, 3 \dots n$ . Since  $Y$ 's are assumed to be independent, the joint probability function or likelihood function is given by:

$$L=L(y; \beta)= \prod_{i=1}^n \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

$L$  is the likelihood function of the parameters. The maximum likelihood principle states that the estimates of the parameters maximize the likelihood function  $L(\beta)$ . However, we take the log of the likelihood function which is called log likelihood and differentiate it with respect to the  $\beta$ 's to get the estimates.

$$\begin{aligned} \text{Log likelihood} = LL = \log L(\beta) &= \sum_{i=1}^n y_i(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k) - \sum_{i=1}^n \log \{1 + \\ \exp \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k\} &- \sum_{i=1}^n \log y_i \end{aligned} \quad (3.6)$$

Differentiating (3.6) with respect to  $\beta$  and equating to zero, we will get the estimates of  $\beta$ .

Manipulating this will be tiresome because of its complexity. But there are statistical soft-wares, such as SAS, STATA and MLWIN, available to solve this iteratively based on Newton- Raphson iterative method.

### 3.3.1.3 Goodness of fit of the model

After estimating the coefficients of the proposed model our first look at the fitted model commonly concerns with an assessment of the goodness of fit of the model and the

significance of the variables in the model .This usually involves formulation and testing of a statistical hypothesis to determine whether the independent variables in the model are "significantly" related to the outcome variable (Hosmer-Lemeshow, 2000).

Goodness of fit of the model can be assessed in two ways first by testing the overall goodness of the model and second by testing the significance of each explanatory variable in the model.

### **The Hosmer–Lemeshow test**

The Hosmer–Lemeshow test is a commonly used test for assessing the goodness of fit of the model and allows for any number of explanatory variables, which may be continuous or categorical. Hosmer and Lemeshow's goodness of fit test divides subjects into  $g$  groups (usually  $g=10$ ) based on predicted probabilities and then computes a chi-square from observed and expected frequencies. Using this grouping strategy, the Hosmer-Lemeshow goodness-of-fit statistic,  $\hat{C}$  is obtained by calculating the Pearson chi-square statistic from the  $g \times 2$  Table of observed and estimated expected frequencies. A formula defining the calculation of  $\hat{C}$  is as follows:

$$\hat{C} = \sum_j^g \frac{(O_j - E_j)^2}{V_j} \quad (3.7)$$

Where  $E_j = n\pi_j$ ,  $V_j = n\pi_j(1-\pi_j)$ ,  $g$  is the number of group,  $O_j$  is observed number of events in the  $j^{th}$  group,  $E_j$  is expected number of events in the  $j^{th}$  group, and  $V_j$  is a variance correction factor for the  $j^{th}$  group. If the observed number of events differs from what is expected by the model, the statistic  $\hat{C}$  will be large and there will be evidence against the null hypothesis that the model is adequate to fit the data. This statistic has an approximate chi-square distribution with  $(g-2)$  degrees of freedom (Agresti, 2007).

If the calculated p-value of the Hosmer-Lemeshow goodness-of-fit test statistic is greater than 0.05, we will not reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the model estimates are adequate to fit the data at an acceptable level.

## Likelihood ratio test

We use likelihood ratio test for checking the significance of the overall model and the significance of individual parameters in the model. Likelihood ratio test is based on comparing the log-likelihood value of two different models; one with  $k$  explanatory variables sometimes called saturated model and the other with  $j$  ( $j < k$ ) explanatory variables. In most cases the comparison is made between the log likelihood of the saturated model and the log likelihood of intercept only (in the absence of explanatory variables) model. If this difference is statistically significant, then the less restrictive model (the one with more variables) is said to fit the data significantly better than the more restrictive model. The likelihood ratio test statistic is given as follows:

$$G^2 = [(-2\ln L_0) - (-2\ln L_1)] \quad (3.8)$$

Where,  $L_0$  is the likelihood of the null model or constant only model and  $L_1$  is the likelihood of the saturated model or the model containing  $k$  predictors.

Under the global null hypothesis,  $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ , the statistic  $G^2$  follows a chi-square distribution with  $k-1$  degrees of freedom and measures how well the independent variables affect the response variable ( Hosmer and Lemeshow, 2000).



**Wald test:** We use the Wald test for testing the significance of individual parameters in logistic regression. That is, the Wald test is used to test

$$H_0: \beta_i = 0 \text{ against } H_1: \beta_i \neq 0, i = 1 \dots k$$

The Wald test is based on the behavior of the log-likelihood function at the ML estimate  $\hat{\beta}$ , having chi-squared form. The Wald test statistic for testing the above hypothesis is:

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)^2}, i = 1, 2, 3 \dots k \quad (3.9)$$

The SE of  $\hat{\beta}$  depends on the curvature of the log-likelihood function at the point where it is maximized, with greater curvature giving smaller SE values.

Another test that we can also use to test the significance of individual parameters is the score test. The score test is based on the behavior of the log-likelihood function at the null value for  $\beta$  of 0. It uses the size of the derivative (slope) of the log-likelihood function, evaluated at the null hypothesis value of the parameter. The derivative at  $\beta=0$  tends to be larger in absolute value when  $\hat{\beta}$  is further from the null value. The score statistic also has an approximate chi-squared distribution with  $df=1$  (Agresti, 2007).

### 3.3.1.4 Residuals for logit models

With categorical predictors, we can use residuals to compare observed and fitted counts. This should be done with the grouped form of the data. Let  $y_i$  denote the number of “successes” for  $n_i$  trials at setting  $i$  of the explanatory variables. Let  $\hat{\pi}_i$  denote the

estimated probability of success for the model fit. Then, the estimated binomial mean  $n_i \hat{\pi}_i$  is the fitted number of successes.

For the logit model, the Pearson residual comparing  $y_i$  to its fit is

$$\text{Pearson residual} = e_i = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{(n_i \hat{\pi}_i [1 - \hat{\pi}_i])}} \quad (3.10)$$

The standard residual divides  $(y_i - n_i \hat{\pi}_i)$  by its SE

$$\text{Standard residual} = \frac{y_i - n_i \hat{\pi}_i}{SE} = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i [1 - \hat{\pi}_i] [1 - h_i]}} \quad (3.11)$$

The term  $h_i$  in this formula is the observation's leverage, its element from the diagonal of the so-called hat matrix. (Roughly speaking, the hat matrix is a matrix that, when applied to the sample logits, yields the predicted logit values for the model.) The greater an observation's leverage ( $h_i > 1/n$  or  $2p/n$  where  $p=k+1$ ), the greater its potential influence on the model fit.

The standardized residual,  $e_i / \sqrt{1 - h_i}$  is larger in absolute value than the Pearson residual,  $e_i$  and is approximately standard normal when the model holds. An absolute value larger than roughly 2 or 3 provides evidence of lack of fit (Agresti, 2007).

### 3.3.1.5 Model diagnostic

Diagnostics are certain quantities computed from the data with the purpose of pinpointing influential points after which these influential points can be removed or corrected.

Influence statistics tell us how much some feature of the model changes when a particular observation is deleted from the model fit.

**DFBETAS**- assess the effect of an individual observation on the estimated parameter of the fitted model. A DFBETAS diagnostic is computed for each observation for each parameter estimate. It is the standardized difference in the parameter estimate due to deleting the corresponding observation. The DFBETAS are useful in detecting observations that causes instability in the selected coefficients.

**Leverage (hat matrix)** value measures the influence of a point on the fit of regression. The centered leverage ranges from 0 (no influence on the fit) to  $(n-1)/n$  or it is from 0 to 1.

**Cook's distance (D)** is a measure of how much the residual of all cases would change if a particular case were excluded from the calculation of the regression coefficients. A large Cook's distance indicates that excluding a case from computation of the regression statistics changes the coefficients substantially (Cook and Weisberg, 1982).

### 3.3.2 Overview of multilevel model

In multilevel research, the structure of the data in the population is hierarchical, and a sample from such a population can be viewed as a multistage sample. First we take a sample of units from the higher level (in our case from the 9 regions and two city administrations), and next we sample the sub units from available units (in our case from mothers). In such samples the individual observations are generally not completely independent.

Because of cost, time and efficiency considerations, stratified multistage samples are the norm for sociological and demographic surveys. For such samples the clustering of the data is, in the phase of data analysis and data reporting, a nuisance which should be taken into consideration.

This clustering sampling scheme often introduces multilevel dependency or correlation among the observations that can have implications for model parameter estimates. For multistage clustered samples, the dependence among observations often comes from several levels of the hierarchy. The problem of dependencies between individual observations also occurs in survey research, where the sample is not taken randomly but cluster sampling from geographical areas is used instead. In this case, the use of single-level statistical models is no longer valid and reasonable. Hence, in order to draw appropriate inferences and conclusions from multistage stratified clustered survey data, we may require tricky and complicated modeling techniques like multilevel modeling (Khan and Shaw, 2011).

The full multilevel regression model assumes that there is a hierarchical data set with one single dependent variable that is measured at the lowest level and explanatory variables at all existing levels. Conceptually the model can be viewed as a hierarchical system of regression equation.

The multilevel model for univariate case with two levels is given by:

$$Y_{ij} = \beta_0 + \beta_{1j}X_{ij} + \mu_{0j} + e_{ij}, j=1,2,\dots,J$$

where: J is number of groups.

$i = 1, 2, \dots, N_i$  Where  $N_i$  number of individuals in each group.

$\mu_{0j}$ =the error at level two.

$e_{ij}$ =the error at level one.

$y_{ij}$  is the response at  $i^{\text{th}}$  level one and  $j^{\text{th}}$  level two

$X_{ij}$  is the explanatory variable at  $i^{\text{th}}$  level one and  $j^{\text{th}}$  level two

The 2011 EDHS data set used for this study is based on stratified, two-stage cluster design, and EAs (sub units of kebele defined in 2007 census) were the sampling units for the first stage. The sample included 624 EAs, 187 in urban areas and 437 in rural areas. Households comprised the second stage of sampling.

Among the appropriate approaches to analyzing under-five child mortality data from this survey is therefore based on nested sources of variability. Here the units at lower level (level-1) are individuals (mothers aged 15–49) who are nested within units at higher level (regions: level-2).

Mothers in each region share the same characteristic and they are not independent.

The response variable in this study, whether mothers have experienced or not experienced under-five mortality during the last five years (2005-2011) is binary.

Consequently multilevel logistic regression is one choice for the analysis. 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 (cluster and division) and individual level variables on individual level outcomes while accounting for the non independence of observations within groups. Also this analysis allows the examination of

both between group and within group variability as well as how group level and individual level variables are related to variability at both levels (Khan and Shaw, 2011).

### 3.3.2.1 Multilevel logistic regression

As described above the data set used for this study has a two level hierarchical structure; mothers (children) are nested within regions. We first consider a two-level model for binary outcomes with a single explanatory variable. Suppose we have data consisting of mothers who gave birth (level one) grouped into Regions (level two). We observe  $Y_{ij}$ , a binary response for mother  $i$  in region  $j$  and  $X_{ij}$ , an explanatory variable at the mother level. Where  $i=1, 2, 3, \dots, n_j$  and  $j=1, 2, 3, \dots, N$ . We define the probability of the response equal to one as  $\pi_{ij} = \Pr(Y_{ij}=1)$  and let  $\pi_{ij}$  be modeled using a logit link function. The standard assumption is that  $Y_{ij}$  has a Bernoulli distribution.

Then the two -level model can be written:

$$\text{logit}(\pi_{ij}) = \log \left[ \frac{\pi_{ij}}{1-\pi_{ij}} \right] = \beta_0 + \beta_1 x_{ij} + U_{0j}, \quad (3.12)$$

Where  $U_{0j}$  is the random effect at level two. Without  $U_{0j}$ , equation (3.12) would be 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  $N(0, \sigma_u^2)$ . Model (3.12) is often described alternatively in the literature as multilevel models as:

$$\log \left[ \frac{\pi_{ij}}{1-\pi_{ij}} \right] = \beta_{0j} + \beta_1 x_{ij} \quad (\text{level one model}) \quad (3.13)$$

$$\text{and } \beta_{0j} = \beta_0 + U_{0j} \quad (\text{level two model}) \quad (3.14)$$

Relative to equation (3.13) and equation (3.14) equation (3.12) is the so-called combined equation (Guang and Zhao, 2000).

The basic data structure of two -level logistic regression is a collection of  $N$  groups (units at level-two(regions))and within group  $j(j=1,2,\dots, N)$  a random sample of  $n_j$  level one(mothers who gave birth)units. The outcome variable is dichotomous and denoted by  $Y_{ij}$  ( $i=1,2,\dots,n_j$ ,  $j=1,2,\dots,N$ ) for level-one unit  $i$  in group  $j$ . The outcomes are supposed to be coded 0 and 1: 0 for ‘failure’, 1 for ‘success’ or vice versa. The total sample size is  $M=\sum_{j=1}^N n_j$ . If one does not take explanatory variables into account, the probability of success is assumed constant in each group. Let the success probability in group  $j$  be denoted by  $P_j$ . The dichotomous outcome variable for the individual  $i$  in group  $j$ ,  $Y_{ij}$ ; which is either 0 or 1 can be expressed as the sum of the probability in group  $j$ ,  $P_j$  (the average proportion of  $j$  levels in group  $j$ ,  $E(Y_{ij}) = P_j$ ) plus some individual-dependent residual  $\varepsilon_{ij}$ . That is  $Y_{ij} = P_j + \varepsilon_{ij}$ . The residual term is assumed to have mean zero and variance,  $\text{Var}(\varepsilon_{ij})= P_j (1-P_j)$ . Since the outcome variable is coded 0 and 1, the group sample average is the proportion of successes in group  $j$  given by  $\hat{P}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}$   $\hat{P}_j$  is an estimate for the group-dependent probability  $P_j$ . Similarly, the overall sample average is the overall proportion of successes,  $\hat{p}$ , and is given by  $\hat{p} = \frac{1}{M} \sum_{j=1}^N \sum_{i=1}^{n_j} Y_{ij}$ . This is an estimate for the overall probability of success,  $p$  (Snijders and Bosker, 1999).

### 3.3.2.1.1 Testing heterogeneous proportions

For the proper application of multilevel analysis in general and multilevel logistic regression analysis in particular, the first logical step is to test for heterogeneity of proportions between groups (in our case between Regions). Two commonly used test statistics that are used to check for heterogeneity of proportions are described below.

To test whether there are indeed systematic differences between groups, the well-known chi-square test for contingency tables can be used. The test statistic of the chi-squared test for a contingency table is often given in the familiar form  $\sum (O - E)^2 / E$  where  $O$  is the observed and  $E$  the expected count in a cell of the contingency table. In this case it can be written also as

$$x^2 = \sum_{j=1}^N n_j \frac{(\hat{p}_{\cdot j} - \hat{p}_{\cdot})^2}{\hat{p}_{\cdot}(1 - \hat{p}_{\cdot})}. \quad (3.15)$$

Where  $\hat{p}_{\cdot j}$  and  $\hat{p}_{\cdot}$  are as given above,  $N$  is the number of groups,  $n_j$  is the number of samples in the  $j^{\text{th}}$  group.

This statistic follows approximately chi-square distribution with  $N-1$  degrees of freedom. The approximation is valid if the expected numbers of success and of failures in each group,  $n_j \hat{p}_{\cdot j}$  and  $n_j(1 - \hat{p}_{\cdot j})$ , respectively, are at least 1 while 80 percent of them are at least 5 (Agresti, 2007). This condition will not always be satisfied, and the chi-square test then may be seriously in error. For a large number of groups the null distribution of the test statistic of the chi-square can be approximated by a normal distribution with the correct mean and variance (McCullagh and Nelder, 1989).



A second test of heterogeneity of proportions was proposed by Commenges and Jacqmin (1994). The proposed test statistic is

$$Z = \frac{\sum_{j=1}^N \{n_j^2 (\hat{p}_j - \hat{p}.)^2\} - M \hat{p}.(1 - \hat{p}.)}{\hat{p}.(1 - \hat{p}.) \sqrt{2 \sum_{j=1}^N n_j (n_j - 1)}} \quad (3.16)$$

The statistic,  $Z$ , follows a standard normal distribution for large values of  $M$  under the null hypothesis. Thus, large calculated values of this statistic are indication of heterogeneous proportions. In the statistic  $Z$  the numerator contains a weight of  $n_j^2$  while chi-square test uses a weight of  $n_j$ . This shows that the two tests combine the groups in different ways. Hence, when the group sizes  $n_j$  are different, it is possible that the two tests lead to different outcomes. The test statistic  $Z$  is shown to have high power over the chi-square test and it can be applied whenever there are many groups, even with small group sizes, provided that no single group dominates. A rule of thumb for the application of this test is that there should be at least  $N=10$  groups, the biggest group should not have a relative larger than  $\frac{n_j}{M} = 0.10$  and the ratio of the largest group size to the 10<sup>th</sup> largest group size should not be more than 10 (Snijders and Bosker, 1999).

**Estimation of between and within -groups variance:** the theoretical variance between the group dependent probabilities, i.e., the population value of  $var(P_j)$ , can be

estimated by:  $\hat{\tau}^2 = S_{between}^2 - \frac{S_{within}^2}{\tilde{n}}$

Where

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

Between- groups variance:

$$S^2_{\text{between}} = \frac{\hat{p}(1-\hat{p})}{\tilde{n}(N-1)} \chi^2, \quad \text{Where } \chi^2 \text{ is as given by equation (3.15), and}$$

Within-groups variance:  $S^2_{\text{with}} = \frac{1}{M-N} \sum n_j p_j (1-p_j).$

### 3.3.2.1.2 The empty logistic regression model

The empty two-level model also called null two-level model for a dichotomous outcome variable refers to a population of groups (level-two units i.e regions) and specifies the probability distribution for group-dependent probabilities  $p_j$  in  $Y_{ij} = p_j + \varepsilon_{ij}$  without taking further explanatory variables into account. We focus on the model that specifies the transformed probabilities  $f(p_j)$  to have a normal distribution. The general link function  $f(p_i)$  is:

$$f(p_i) = \beta_0 + U_{0j} \quad (3.17)$$

where,  $\beta_0$  is the population average of the transformed probabilities and  $U_{0j}$  the random deviation from this average for group  $j$ . If  $f(p)$  is the *logit* function, then  $f(p_j)$  is just

the log-odds for group  $j$ . Thus, for the *logit* link function, the log-odds have a normal distribution in the population of groups, and are given by:

$$\log \text{it}(p_j) = \beta_0 + U_{0j} \quad (3.18)$$

$$U_{0j} \sim \text{IID}(0, \sigma_u^2)$$

This model does not include a separate parameter for the level-one variance. This is because the level-one residual variance of the dichotomous outcome variable follows directly from the success probability, as indicated by equation  $\text{var}(\epsilon_{ij}) = p_j(1-p_j)$ . The probability corresponding to the average value  $\beta_0$ , denoted by  $\pi_0$ , is defined by  $f(\pi_0) = \beta_0$ .

For the logit function, the so-called logistic transformation of  $\beta_0$ , is defined by

$$\pi_0 = \text{logit}(\beta_0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \quad (\text{Snijders and Bosker, 1999}). \quad (3.18)$$

The null model also serves as a “baseline model” for purposes of comparison with later, more complex models. Note that a model is “conditional” by the presence of predictors at level 1 or level 2. Since the researcher almost always employs predictor variables and is not simply interested in the null model, most mixed models are conditional.

### 3.3.2.1.3 The random intercept logistic regression

Random intercept models are models where only the intercept of the level 1 dependent variable is modeled as an effect of the level 2 grouping variable and possibly other level 1 or level 2 (or higher). Random intercept regression models are also called “means-as-outcomes regression models.” This variant of the random intercept model predicts the level 1 intercept on the basis of the level 2 grouping variable and also on the basis of one or more level 2 random effect predictors. Random intercepts are used to model unobserved heterogeneity in the overall response. This allows the overall probability of infant and child mortality to vary across Regions.

In the random intercept logistic regression model, the intercept is the only random effect meaning that the groups differ with respect to the average value of the response variable.

It represents the heterogeneity between groups in the overall response. The logistic random intercept model expresses the log-odds, i.e. the logit of  $P_{ij}$ , as a sum of a linear function of the explanatory variables and a random group-dependent deviation  $U_{0j}$ . That is,

$$\begin{aligned}
 \log \text{it}(p_{ij}) &= \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} \dots + \beta_k x_{kij} \\
 &= \beta_{0j} + \sum_{h=1}^k \beta_h x_{hij} \\
 &= \beta_0 + \sum_{h=1}^k \beta_h x_{hij} + U_{0j}
 \end{aligned} \tag{3.19}$$

where the intercept term  $\beta_{0j}$  is assumed to vary randomly and is given by the sum of an average intercept  $\beta_0$  and group-dependent deviations,  $u_{0j}$ . That is  $\beta_{0j} = \beta_0 + U_{0j}$ . The

first part incorporating the regression coefficients  $\beta_0 + \sum_{h=1}^k \beta_h x_{hj}$  is the fixed part of the model, because the coefficients are fixed. The remaining part  $u_{0j}$  is called the random part of the model. It is assumed that the residual,  $U_{0j}$  are mutually independent and normally distributed with mean zero and variance  $\sigma_0^2$  (Snijders and Bosker, 1999).

As a result

$$\text{logit}(p_{ij}) = \beta_0 + \sum_{h=1}^k \beta_h x_{hj} + U_{0j} \quad (3.20)$$

Solving for  $P_{ij}$  we obtain

$$P_{ij} = \frac{e^{\beta_0 + \sum_{h=1}^k \beta_h x_{hj} + U_{0j}}}{1 + e^{\beta_0 + \sum_{h=1}^k \beta_h x_{hj} + U_{0j}}} \quad (3.21)$$

Thus, a unit difference between the  $x_h$  values of two individuals in the same group is associated with a difference of  $\beta_h$  in their log-odds, or equivalently, a ratio of  $\exp(\beta_h)$  which is the odds ratio, comparing odds for individuals in the same group.  $U_{0j}$  is the effect of being in group  $j$  on the log -odds that  $y = 1$  also known as a level 2 residual,  $\sigma_0^2$  is the level 2 (residual) variance, or the between-group variance in the log-odds that  $y = 1$  after accounting for  $X$ .

Note that the first part of the left-hand side of (3.20), incorporating the regression coefficients, is the fixed part of the model, because the coefficients are fixed. The remaining part,  $U_{0j}$  is called the random part of the model.

Random intercept models have many applications, for instance estimating the regional effects on under-five child mortality, adjusting for individual mother's level factors, and within the model, evaluate and compare the performance of the region's infant and child mortality reduction. This can be done by obtaining the odds ratio for each region. This regional effect is a measure of the situation of infant and child mortality due to the region relative to the average of all regions. If the odd of under-five child mortality for regional effects is sufficiently larger than one, the region is considered to have performed worse than the average; if it is significantly smaller than one, the region is considered to have better performance than the expected (DeLong et al., 1997).

#### **3.3.2.1.4 The Random Coefficient multilevel Logistic Regression Model**

Random coefficients models are ones where the coefficient(s) of lower-level predictor(s) is/are modeled as well. Random coefficients model unobserved heterogeneity in the effects of explanatory variables on the response variable.

In logistic regression analysis, linear models are constructed for the log-odds. The multilevel analogue, random coefficient logistic regression, is based on linear models for the log-odds that include random effects for the groups or other higher level units.

Consider explanatory variables which are potential explanations for the observed outcomes.

Denote these variables by  $X_1, X_2, \dots, X_k$ . The values of  $X_h (h=1, 2, \dots, k)$  are indicated in the usual way by  $x_{hij}$ . Since some or all of these variables could be level-one variable, the success probability is not necessarily the same for all individuals in a given group. Therefore, the success now consider a model with group-specific regressions of logit of the success probability,  $\text{logit}(P_{ij})$ , on a single level one explanatory variable  $X$

$$\text{logit}(P_{ij}) = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} \quad (3.22)$$

The intercepts  $\beta_{0j}$  as well as the regression coefficients, or slopes,  $\beta_{1j}$  are group-dependent.

These group-dependent coefficients can be split into an average coefficient and the group dependent deviation:

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

$$\beta_{1j} = \beta_1 + U_{1j}$$

Substituting into equation (3.22) leads to the model

$$\begin{aligned} \log \text{it}(p_{ij}) &= \log \left[ \frac{p_{ij}}{1 - p_{ij}} \right] = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})x_{1ij} \\ &= \beta_0 + \beta_1 x_{1ij} + U_{0j} + U_{1j}x_{1ij} \end{aligned} \quad (3.23)$$

There are two random group effects, the random intercept  $U_{0j}$  and the random slope  $U_{1j}$ . It is assumed that the level-two residuals  $U_{0j}$  and  $U_{1j}$  have means zero given the value of the explanatory variable  $X$ . Thus  $\beta_1$  is the average regression coefficient like  $\beta_0$  is the average intercept. The first part of equation (3.23),  $\beta_0 + \beta_1 x_{1ij}$  is called the fixed part of the model and the second part,  $U_{0j} + U_{1j}x_{1ij}$  is called the random part. The term  $U_{1j}x_{1ij}$  can be regarded as a random interaction between group and  $X$ . This model implies that the groups are characterized by two random effects: their intercept and their slope. These two group effects,  $U_{0j}$  and  $U_{1j}$  will not be independent, but correlated. Further, it is assumed that, for different groups, the pairs of random effects  $(U_{0j}, U_{1j})$  are independent and

identically distributed. The random intercept variance,  $\text{var}(U_{0j}) = \sigma_0^2$ , the random slope variance,  $\text{var}(U_{1j}) = \sigma_1^2$ , and the covariance between the two random effects  $\text{Covar}(U_{0j}, U_{1j}) = \sigma_{01}$ , are called variance components (Snijders and Bosker, 1999).

The model for a single explanatory variable discussed above can be extended by including more variables that have random effects. Suppose that there are  $k$  level-one explanatory variables  $X_1, X_2, \dots, X_k$ , and consider the model where all  $X$ -variables have varying slopes and random intercept. That is,

$$\text{logit}(p_{ij}) = \log \left[ \frac{p_{ij}}{1-p_{ij}} \right] = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \dots + \beta_{kj}x_{kij} \quad (3.24)$$

Letting

$$\beta_{0j} = \beta_0 + U_{0j} \quad \text{and} \quad \beta_{hj} = \beta_h + u_{hj}, \quad \text{for } h = 1, 2, \dots, k$$

$$\text{logit}(p_{ij}) = \log \left[ \frac{p_{ij}}{1-p_{ij}} \right] = \beta_0 + \sum_{h=1}^k \beta_h X_{hij} + U_{0j} + \sum_{h=1}^k U_{hj} X_{hij} \quad (3.25)$$

The first part of this model,  $\beta_0 + \sum_{h=1}^k \beta_h X_{hij}$  is the fixed part and the second part  $U_{0j} + \sum_{h=1}^k U_{hj} X_{hij}$ , is the random part of the model (Snijders and Boskers, 1999)

### 3.3.2.1.5 Intra-class Correlation Coefficient (ICC)



The fundamental reason of using multilevel analysis is the existence of more similarities between mothers in the same Region than in different Regions. This leads to the existence of intra-class (intra-Regional) correlation of infant and child mortality.

ICC is the degree of resemblance between level one units belonging to the same group. It is an indication of the proportion of variance at the second level (region) and it can also be interpreted as the expected (population) correlation between two randomly chosen individuals within the same group (Joop ,2010)

In two- level model, the ICC is calculated in the intercept only model. This model can be derived from “Eq. (3.23)” by excluding all explanatory variables, which results in the following equation:  $\log it(p_j) = \beta_0 + U_{0j}$ . The ICC is then calculated based on the following formula:

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} \quad (3.24)$$

where,  $\sigma_e^2$  is variance of individual (lower) level units.

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

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

### **3.3.2.1.6 Parameter estimation**

Parameter estimation for multilevel logistic model is not straightforward like the methods for ordinary logistic regression. The most common methods for estimating multilevel logistic models are based on likelihood. Among the methods, Marginal Quasi Likelihood or MQL [Goldstein (1991), Goldstein and Rasbash (1996)] and Penalized Quasi Likelihood or PQL [Laird (1978) and Breslow and Clayton (1993)] are the two prevailing approximation procedures. Both MQL and PQL are based on Taylor series expansion to achieve the approximation. Based on the first and second term of Taylor expansion, MQL and PQL are often known as first order MQL and second-order MQL, first-order PQL and second-order MQL respectively. After applying these quasi likelihood methods, the model is then estimated using iterative generalized least squares (IGLS) or reweighted IGLS (RIGLS) [Goldstein (2003)].

Besides, there are other estimation methods: Maximum Likelihood Method (several simulation based; McCulloch (1997)), Bayesian methods using Markov Chain Monte Carlo (MCMC), adaptive Gaussian quadrature (AGQ) and the Iterative Bootstrap method. Using MCMC simulation technique has come to the forefront of statistical research over the last one and half decade [Gelfand et al. (1990)] and also it is being used with greater extent in multilevel modeling recently.

### **3.3.2.1.7 Significance testing**

As with ordinary least squares regression or logistic regression, we can consider significance tests for individual estimates, such as intercepts, slopes, and their variances, as well as whether the full model accounts for a significant amount of variance in the

dependent variable. In between, there is also the possibility of determining whether the subset of predictors contribute significantly.

The fixed effects in multilevel regression are typically tested in a familiar way, by creating a ratio of the intercept or slope estimate to the estimate of the standard error. The usual null hypothesis test is whether the coefficient, either intercept or slope, is significantly different from zero (i.e., is the population value zero or not). This kind of ratio, usually distributed as a z or t, is used in many statistical tests (Newsom, 2011).

$$t = \frac{\hat{\beta}_h}{SE(\hat{\beta}_h)}$$

Where  $\hat{\beta}_h$  is either the intercept or slope coefficient and S.E ( $\hat{\beta}_h$ ) is the standard error estimate.

In SPSS (and similarly in SAS, Stata, and MLwin), fixed effects tests involve the same ratio of the estimate to the standard error estimate, but significance is determined by the normal curve, so it is considered a z-test. The z-test is often referred to as a “Wald” test.

### **Significance Testing for Random Effects**

Random effects tests examine hypotheses about whether the variance of intercept or slopes (or their covariance) is significantly different from zero. The tests of variances and covariances are made using a Wald z-test and chi-square test. The Wald test for variances is simply a ratio of the variance estimate divided by the standard error estimate.

Significance tests of variances (but not covariance) using this approach should be interpreted after dividing the p-Value from the output in half (i.e., as a one-tailed test; Snijders & Bosker, 1999).

### **Likelihood Ratio Tests for Multiple Parameters**

Another approach to significance tests involves a comparison of two “nested models.” Nested model tests involve comparison of one model to another model that specifies only a subset of the parameters included in the first model (provided the same set of cases are used in both models).

The likelihood ratio test compares the deviance ( $-2 \log$  likelihood) of two models by subtracting the smaller deviance (model with more parameters) from the larger deviance (model with larger deviance). The difference is a chi-square test with the number of degrees of freedom equal to the number of different parameters in the two models. Any number of parameters can be compared in the two models. In the case where the empty model is compared to a full model, the likelihood ratio test provides information about whether the predictors in the model together account for a significant amount of variance in the dependent variable.

Similarly, the overall model evaluation is also examined using Akaike Information Criteria (AIC) and Schwartz Information Criteria (BIC). The smaller the value, the better the model will be.

$$AIC = -2 \times \ln(\text{likelihood}) + 2 \times k$$

$$BIC = -2 \times \ln(\text{likelihood}) + \ln(N) \times k$$

where  $k$  is the model degrees of freedom calculated as the rank of variance–covariance matrix of the parameters and  $N$  is the number of observations used in estimation or, more precisely, the number of independent terms in the likelihood.

## **Chapter four**

### **4. Statistical data analysis and results**

The objective of this chapter is to provide analysis of results about the effect of socioeconomic and demographic characteristics on under-five child mortality. The analysis was done using SPSS version 17 and STATA version 11.

This chapter is divided into four sections. The first section presents the results of descriptive analysis; the second section, the results of cross tabulation analysis; the third section, analysis of the results obtained from multiple logistic regressions; and the last section, analysis of the results obtained from multilevel logistic regression.

#### **4.1 Results of descriptive analysis**

The distribution of socioeconomic and demographic background characteristics of mothers which affect under-five child mortality is presented in Table 4.1 below. A total of 11,250 mothers having live births were considered in this study. From these, 814 (7.24%) have lost their under five children and 10,445 (92.84%) of the mothers had children survived at the time of the survey.

Table 4.1 Results of descriptive analysis of socio-economic and demographic factors.

| Variables                   | Categories             | Under-five mortality |            |       |            |       |
|-----------------------------|------------------------|----------------------|------------|-------|------------|-------|
|                             |                        | Yes                  | Percentage | NO    | Percentage | Total |
|                             | Tigray                 | 78                   | 6.75       | 1,078 | 93.25      | 1,156 |
|                             | Afar                   | 95                   | 8.65       | 1,003 | 91.35      | 1,098 |
|                             | Amhara                 | 88                   | 7.03       | 1,163 | 92.97      | 1,251 |
|                             | Oromia                 | 122                  | 7.06       | 1,607 | 92.94      | 1,729 |
| Region                      | Somali                 | 73                   | 7.45       | 907   | 92.55      | 980   |
|                             | Benishangul-Gumuz      | 91                   | 9.15       | 903   | 90.85      | 994   |
|                             | SNNP                   | 116                  | 7.41       | 1449  | 92.59      | 1,565 |
|                             | Gambela                | 57                   | 7.18       | 737   | 92.82      | 794   |
|                             | Harari                 | 43                   | 6.71       | 598   | 93.29      | 641   |
|                             | Addis Ababa            | 11                   | 2.96       | 361   | 97.04      | 372   |
|                             | Dire Dawa              | 31                   | 4.63       | 639   | 95.37      | 670   |
| Type of place of residence  |                        |                      |            |       |            |       |
|                             | Urban                  | 108                  | 5.72       | 1,781 | 94.28      | 1,889 |
|                             | Rural                  | 697                  | 7.45       | 8,664 | 92.55      | 9,361 |
| Mother's educational level  |                        |                      |            |       |            |       |
|                             | No education           | 589                  | 7.48       | 7,288 | 92.52      | 7,877 |
|                             | Primary                | 196                  | 6.97       | 2,618 | 93.03      | 2,814 |
|                             | Secondary              | 17                   | 4.58       | 354   | 95.42      | 371   |
|                             | Higher                 | 3                    | 1.60       | 185   | 98.40      | 188   |
| Household has television    |                        |                      |            |       |            |       |
|                             | Yes                    | 63                   | 5.74       | 1,034 | 94.26      | 1,097 |
|                             | No                     | 725                  | 7.33       | 9,169 | 92.67      | 9,894 |
|                             | Not a de jure resident | 17                   | 6.56       | 242   | 93.44      | 259   |
| Religion                    |                        |                      |            |       |            |       |
|                             | Orthodox               | 238                  | 6.83       | 3,249 | 93.17      | 3,487 |
|                             | Catholic               | 11                   | 10.38      | 95    | 89.62      | 106   |
|                             | Protestant             | 144                  | 6.75       | 1,990 | 93.25      | 2,134 |
|                             | Muslim                 | 391                  | 7.40       | 4,896 | 92.60      | 5,287 |
|                             | Traditional            | 9                    | 9.38       | 87    | 90.63      | 96    |
|                             | Other                  | 12                   | 8.57       | 128   | 91.43      | 140   |
| Number of household members |                        |                      |            |       |            |       |
|                             | 1-5                    | 457                  | 9.11       | 4,558 | 90.89      | 5,015 |
|                             | 6-10                   | 324                  | 5.61       | 5,448 | 94.39      | 5,772 |
|                             | 11-15                  | 19                   | 4.80       | 377   | 95.20      | 396   |
|                             | 16 and above           | 5                    | 7.46       | 62    | 92.54      | 67    |
| Sex of household head       |                        |                      |            |       |            |       |

|                                   |                           |     |       |       |       |       |
|-----------------------------------|---------------------------|-----|-------|-------|-------|-------|
|                                   | Male                      | 650 | 7.09  | 8,516 | 92.91 | 9,166 |
|                                   | Female                    | 155 | 7.44  | 1,929 | 92.56 | 2,084 |
| Wealth index                      |                           |     |       |       |       |       |
|                                   | Poorest                   | 287 | 8.20  | 3,213 | 91.80 | 3,500 |
|                                   | Poorer                    | 155 | 7.56  | 1,895 | 92.44 | 2,050 |
|                                   | Middle                    | 127 | 6.99  | 1,691 | 93.01 | 1,818 |
|                                   | Richer                    | 117 | 6.48  | 1,688 | 93.52 | 1,805 |
|                                   | Richest                   | 119 | 5.73  | 1,958 | 94.27 | 2,077 |
| Total children ever born          |                           |     |       |       |       |       |
|                                   | 1-4                       | 431 | 6.50  | 6,202 | 93.50 | 6,633 |
|                                   | 5-8                       | 289 | 7.65  | 3,490 | 92.35 | 3,779 |
|                                   | 9 and above               | 85  | 10.14 | 753   | 89.86 |       |
| Age of mother a first birth       |                           |     |       |       |       |       |
|                                   | 9-14                      | 69  | 7.99  | 795   | 92.01 | 864   |
|                                   | 15-19                     | 462 | 7.61  | 5,606 | 92.39 | 6,068 |
|                                   | 20-24                     | 225 | 6.49  | 3,244 | 93.51 | 3,469 |
|                                   | 25-29                     | 36  | 5.23  | 652   | 94.77 | 688   |
|                                   | 30 and above              | 13  | 8.07  | 148   | 91.93 | 161   |
| Current marital status            |                           |     |       |       |       |       |
|                                   | Married                   | 723 | 7.31  | 9,174 | 92.69 | 9,897 |
|                                   | Living with partner       | 32  | 5.64  | 535   | 94.36 | 567   |
|                                   | Widowed                   | 15  | 7.25  | 192   | 92.75 | 207   |
|                                   | Divorced                  | 29  | 7.40  | 363   | 92.60 |       |
|                                   | No longer living together | 6   | 3.21  | 181   | 96.79 | 187   |
| Husband/partner's Education level |                           |     |       |       |       |       |
|                                   | No education              | 453 | 7.79  | 5,362 | 92.21 | 5,815 |
|                                   | Primary                   | 274 | 6.71  | 3,808 | 93.29 | 4,082 |
|                                   | Secondary                 | 44  | 5.64  | 736   | 94.36 | 780   |
|                                   | Higher                    | 27  | 5.76  | 442   | 94.24 | 469   |
|                                   | Don't know                | 7   | 6.73  | 97    | 93.27 | 104   |
| Husband/partner's occupation      |                           |     |       |       |       |       |
|                                   | Jobless                   | 116 | 5.81  | 1,880 | 94.19 | 1,996 |
|                                   | Managers                  | 0   | 0.00  | 1     | 100   | 1     |
|                                   | Agricultural worker       | 617 | 7.64  | 7,464 | 92.36 | 92.36 |
|                                   | Trade worker              | 23  | 5.23  | 417   | 94.77 | 440   |
|                                   | Operator or assembler     | 9   | 4.27  | 202   | 95.73 | 211   |
|                                   | Laborer                   | 40  | 7.86  | 469   | 92.14 | 509   |
|                                   | Other                     | 0   | 0.00  | 12    | 100   | 12    |
| Mother's occupation               |                           |     |       |       |       |       |
|                                   | Not working               | 430 | 6.99  | 5,721 | 93.01 | 6,151 |
|                                   | Professional/technical    | 5   | 2.91  | 167   | 97.09 | 172   |
|                                   | Sales/services            | 115 | 6.55  | 1,641 | 93.45 | 1,756 |
|                                   | Agricultural              | 186 | 7.92  | 2,163 | 92.08 | 2,349 |
|                                   | Manual                    | 69  | 8.39  | 753   | 91.61 | 822   |

|                          |                                                   |     |       |       |       |       |
|--------------------------|---------------------------------------------------|-----|-------|-------|-------|-------|
| Birth order number       |                                                   |     |       |       |       |       |
|                          | 1-3                                               | 407 | 7.04  | 5,378 | 92.96 | 5,785 |
|                          | 4-6                                               | 255 | 7.16  | 3,307 | 92.84 | 3,562 |
|                          | 7-9                                               | 120 | 7.74  | 1,431 | 92.26 | 1,551 |
|                          | 10 and above                                      | 23  | 6.53  | 329   | 93.47 | 352   |
| Sex of child             |                                                   |     |       |       |       |       |
|                          | Male                                              | 442 | 7.64  | 5,340 | 92.36 | 5,782 |
|                          | Female                                            | 363 | 6.64  | 5,105 | 93.36 | 5,468 |
| Months of breast feeding |                                                   |     |       |       |       |       |
|                          | 0-6 months                                        | 1   | 0.07  | 1,336 | 99.93 | 99.93 |
|                          | 7-12 months                                       | 9   | 0.96  | 938   | 100   | 938   |
|                          | 13-18 months                                      | 2   | 0.23  | 868   | 99.77 | 870   |
|                          | 19-59 months                                      | 3   | 0.20  | 1,486 | 99.80 | 1,489 |
|                          | Ever breast feeding, not currently breast feeding | 542 | 8.76  | 5,646 | 91.24 | 6,188 |
|                          | Never breast fed                                  | 257 | 60.05 | 171   | 39.95 | 428   |
| Age of mother            |                                                   |     |       |       |       |       |
|                          | 15-19                                             | 50  | 10.12 | 444   | 89.88 | 494   |
|                          | 20-24                                             | 187 | 8.37  | 2,048 | 91.63 | 2,235 |
|                          | 25-29                                             | 215 | 6.09  | 3,313 | 93.91 | 3,528 |
|                          | 30-34                                             | 140 | 6.04  | 2,179 | 93.96 | 2,319 |
|                          | 35-39                                             | 140 | 8.17  | 1,574 | 91.83 | 1,714 |
|                          | 40-44                                             | 50  | 7.00  | 664   | 93.00 | 714   |
|                          | 45-49                                             | 23  | 9.35  | 223   | 90.65 | 246   |

The proportion of under-five child mortality varied from one region to another region with the highest proportion (9.15%) of under-five child mortality recorded in Benishangul-Gumuz followed by Afar (8.6%) and Somali (7.45%). However, the least proportion of under-five child mortality was observed in Addis Ababa (2.96%) followed by Diredawa (4.63%), Harari (6.71%) and Tigray (6.75%). Amhara (7.03%) and Oromia (7.06%) had the same proportion of under-five child mortality which is less than the rate in SNNP (7.41%) but higher than the rate in Tigray.

Under-five child mortality in urban areas was also different from the rate in rural areas. The proportion of under-five child mortality in urban areas was 5.72% while it was



7.45% in rural areas. This shows that the rate of under-five child mortality in rural areas was about 1.3 times the rate in urban areas.

Under-five child mortality was different for different levels of mother's education. About 7.48% of under-five children mortality was recorded for non-educated mothers while 6.97% of the children died before the age of 5 were born to mothers with primary education. About 4.58% of children born to mothers that have completed secondary school could not celebrate their fifth birth day while only 1.6% of children born to mothers that have completed higher education died. On the other hand the proportion of under-five child mortality was 7.79% for those born from those who completed uneducated fathers compared to 6.71% of under-five children born to primary educated fathers. But under-five child mortality was not remarkably different for those born from fathers who completed secondary and higher education with percentages 5.64% and 5.76%, respectively.

Moreover, the rate of under-five child mortality was different among the different religion categories. The highest proportion (10.38%) of under-five child mortality was observed among Catholic followers followed by traditional followers (9.38%) but the least proportion (6.75%) of under-five child mortality was observed among protestant followers followed by orthodox followers (6.83%). Number of household members indicated unexpected variation of under-five child mortality. The highest proportion (9.11%) of U5CM was observed among households of family size 1-5 followed by the households of size 16 and more (7.46%). On the contrary, households with 6-10 members had 5.61% U5CM next to households with 11-15 members which accounted for 4.8% of under-five children death.

The proportion of under-five children death was nearly invariable for both sexes of household head. In contrast, wealth index of parents showed different proportion of under-five child mortality exposure. Children born to poor families had the highest proportion of mortality (8.2% for poorest families and 7.56% for poorer families) but children born to rich families had the least proportion of under-five child mortality (5.73% from richest families and 6.48% from richer families).

The experience of U5CM was high (10.14%) for households with a total of at least 9 children ever born to the mother followed by households with a total of 5-8 children ever born to the mother ( 7.65 %). However U5CM was least (6.5%) for mothers with 1-4 children. The proportion of U5CM was 8% for mothers of age 9-14 years and for mothers aged 30 and above years at first birth. On the other hand, it was 5.23% for mothers in the age group of 25-29 years at first birth followed by 6.49% and 7.61% for mothers aged 20-24 years and 15-19 years respectively.

The risk of under-five child mortality was different for the different categories of marital status. About 7.31% of the children died before the age of 5 were born to married mothers, 7.25% to widowed mothers and 7.4% to divorced mothers. These three categories of marital status had almost the same contribution. About 3.21% of under-five children death was observed with mothers who were not living with their husbands or partners while 5.64% of the under 5 children death was recorded for mothers living with their partners.

While 7.64% of the children died before the age of 5 were from agricultural worker fathers, 7.86% of the children who died were from fathers whose occupation was laborer.

5.81%, 5.23% and 4.27% of under-five children death were observed for households where the fathers were jobless, trade worker and operator or assembler, respectively. Similarly the highest proportion (8.39%) of under-five children mortality was recorded for mothers whose occupation is manual followed by agricultural worker mothers who lost 7.92% of their under-five children which was still higher than the proportion for not working mothers who lost 6.99% of their under-five children. Conversely, only 2.91% of the under-five children mortality was observed for mothers whose occupation was professional.

There was no significant difference of under- five children survival for the different birth orders except the birth order number of 10 and above. The proportion of under-five children mortality for birth order numbers of 1-3, 4-6 and 7-9 were 7.04%, 7.16% and 7.74%, respectively. The percentage of female and male children surviving was 7.64%, 6.64%, respectively.

The highest percentage (60.05%) of never breast fed under-five children died while mothers who ever breast fed lost 8.76% of their children. No under-five children death was observed for mothers who breast fed from 7-12 months followed by 0.07% of under-five children death for mothers who breast fed for 0-6 months. The highest percentage (10.12%) of U5CM was recorded for mothers aged 15-19 years, 9.35% for mothers aged 45-49 followed by 8.37% and 8.17% of U5CM observed for mothers aged 20-24 and 35-39 years, respectively.

## 4.2 Results of cross tabulation data analysis

Cross-tabulation analysis, also known as contingency table analysis, is often used to analyze categorical (nominal or ordinal measurement scale) data. A cross-tabulation is a two (or more) dimensional table that records the number (frequency) of respondents that have the specific characteristics described in the cells of the table.

The chi-square ( $\chi^2$ ) test is used to assess the relationship between two nominal or ordinal variables. It is a very general statistical test that can be used whenever we wish to evaluate whether frequencies that have been empirically obtained differ significantly from those that would be expected on the basis of chance or theoretical expectations.

Chi-square statistic does not give any information about the strength of the relationship and only conveys the existence or nonexistence of the relationships between the variables investigated.

Based on the result of the cross tabulation analysis, under-five child mortality was found to be associated with place of residence ( $p = 0.008$ ), with mother's educational level ( $p=0.003$ ), with number of house hold members( $p=0.000$ ), wealth index ( $p=0.008$ ), husband/partner's occupation ( $p=0.023$ ), mother's occupation ( $p=0.043$ ), sex of child ( $p=0.039$ ), months of breast feeding ( $p=0.000$ ), age of mother ( $p=0.000$ ) and region ( $p=0.003$ ). (see the Appendix).

## 4.3 Logistic regression analysis

### 4.3.1 Estimation and goodness of fit test

Multiple logistic regression analysis was used to identify the influence of each socio-economic and demographic variable on under-five child mortality. The analysis was done by using STATA with stepwise variable selection method.

Table 4.2 shows statistically significant variables and the maximum likelihood estimates of parameters where the significance of each parameter is tested using the Wald test.

Table 4.2 Results of maximum likelihood estimates of parameters in fitting logistic regression

| Variables with categories    | Coefficient | Std. Err | Z      | p> Z   | Odds ratio | [95%Conf.Interval for OR] |        |
|------------------------------|-------------|----------|--------|--------|------------|---------------------------|--------|
| Months of breast feeding     |             |          | 25.12  | .000*  |            |                           |        |
| 0-6 Months                   | -7.858      | 1.007    | -7.8   | 0.000* | .0003      | .000                      | .003   |
| 6-12 months                  | -5.280      | 0.355    | -14.87 | 0.000* | .005       | .002                      | .010   |
| 13-18 months                 | -6.766      | .718     | -9.43  | 0.000* | .001       | .000                      | .005   |
| 19-59 months                 | -6.884      | .590     | -11.68 | 0.000* | .001       | .044                      | .071   |
| Ever breast fed              | -2.877      | .119     | -24.2  | 0.000* | .056       |                           |        |
| Never breast fed(ref)        |             |          |        |        |            | 14.070                    | 22.421 |
| Birth order number           |             |          | 1.46   | 0.006* |            |                           |        |
| 1-3(ref)                     |             |          |        |        |            |                           |        |
| 4-6                          | .541        | .188     | 2.88   | 0.004* | 1.718      | 1.188                     | 2.482  |
| 7-9                          | .603        | .269     | 2.24   | 0.025* | 1.827      | 1.079                     | 3.093  |
| 10 and above                 | -.080       | .411     | -0.190 | 0.846  | .923       | .412                      | 2.067  |
| Age of mother at first birth |             |          | 1.46   | 0.143  |            |                           |        |
| 9-14                         | -.877       | .408     | -2.15  | 0.032* | .416       | .187                      | .926   |
| 15-19                        | -.745       | .373     | -2.00  | 0.046* | .571       | .228                      | .987   |
| 20-24                        | -.637       | .362     | -1.76  | 0.078  | 1.192      | .260                      | 1.074  |
| 25-29                        | -.565       | .387     | -1.46  | 0.144  | .474       | .266                      | 1.213  |
| 30 and above(ref)            |             |          |        |        |            |                           |        |
| Total children ever born     |             |          | 4.32   | 0.000* |            |                           |        |
| 1-4(ref)                     |             |          |        |        |            |                           |        |
| 5-8                          | .418        | .185     | 2.26   | 0.024* | 1.519      | 1.058                     | 2.182  |
| 9 and above                  | 1.210       | .287     | 4.21   | 0.000* | 3.352      | 1.909                     | 5.888  |
| Number of house hold members |             |          | -8.77  | 0.000* |            |                           |        |

|                           |        |       |       |        |       |       |        |
|---------------------------|--------|-------|-------|--------|-------|-------|--------|
| 1-5(ref)                  |        |       |       |        |       |       |        |
| 6-10                      | -.942  | .109  | -8.61 | 0.000* | .390  | .314  | .483   |
| 11-15                     | -1.385 | .279  | -4.96 | 0.000* | .250  | .145  | .433   |
| 16 and above              | -.447  | .525  | -0.85 | 0.394  | .639  | .228  | 1.79   |
| Mothers educational level |        |       | -3.33 | 0.001* |       |       |        |
| No education              | 1.877  | .756  | 2.48  | 0.013* | 6.536 | 1.484 | 28.775 |
| Primary                   | 1.808  | .752  | 2.41  | 0.016* | 6.103 | 1.399 | 26.633 |
| Secondary                 | 1.234  | .750  | 1.65  | 0.100  | 3.436 | .790  | 14.945 |
| Higher(ref)               |        |       |       |        |       |       |        |
| Region                    |        |       | -2.25 | 0.025* |       |       |        |
| Tigray                    | .640   | .391  | 1.64  | 0.101  | 1.896 | .882  | 4.077  |
| Afar                      | 1.079  | .386  | 2.80  | 0.005* | 2.941 | 1.381 | 6.261  |
| Amhara                    | .821   | .388  | 2.12  | 0.034* | 2.273 | 1.063 | 4.861  |
| Oromia                    | .767   | .378  | 2.03  | 0.043* | 2.153 | 1.026 | 4.518  |
| Somali                    | .526   | .392  | 1.34  | 0.180  | 1.691 | .785  | 3.646  |
| Benishangul-Gumuz         | .955   | .385  | 2.48  | 0.013* | 2.600 | 1.221 | 5.534  |
| SNNP                      | 1.039  | .390  | 2.66  | 0.008* | 2.826 | 1.315 | 6.071  |
| Gambela                   | .890   | .416  | 2.14  | 0.032* | 2.436 | 1.078 | 5.505  |
| Harari                    | .658   | .396  | 1.66  | 0.096  | 1.932 | .890  | 4.194  |
| Addis Ababa(ref)          |        |       |       |        |       |       |        |
| Diredawa                  | .396   | .402  | 0.99  | 0.325  | 1.485 | .676  | 3.264  |
| Age of mother             |        |       | -4.60 | 0.000* |       |       |        |
| 15-19                     | 1.529  | .405  | 3.78  | 0.000* | 4.614 | 2.086 | 10.204 |
| 20-24                     | .819   | .357  | 2.29  | 0.022* | 2.267 | 1.127 | 4.562  |
| 25-29                     | .398   | .328  | 1.21  | 0.225  | 1.489 | .782  | 2.833  |
| 30-34                     | .032   | .308  | 0.11  | 0.916  | 1.033 | .565  | 1.887  |
| 35-39                     | .206   | .285  | 0.73  | 0.468  | 1.229 | .704  | 2.148  |
| 40-44                     | -.236  | .299  | -0.79 | 0.429  | .790  | .440  | 1.418  |
| 45-49(ref)                |        |       |       |        |       |       |        |
| Constant                  | -4.736 | 1.121 | -4.23 | 0.000* |       |       |        |

“\*” indicates significance at  $\alpha=0.05$ . Ref=reference

The reference categories are selected on the basis of low frequency of the event; ease of comparison and by default. STATA codes the reference category as 0 and the remaining categories as 1 alternatively keeping others as 0.

In order to check the goodness-of-fit of an estimated multiple logistic regression model one should assume that the model contains those variables that should be in the model and have been entered in the correct functional form.

The likelihood ratio test is used to test the goodness of fit of the model by comparing two nested models:-one with small number of explanatory variables and the other with more explanatory variables. Two models will be compared: one with no variable called empty (intercept only) model and the other with all variables called saturated (full) model.

The null hypothesis is:

$H_0$ : There is no significant difference between the empty model and the saturated model.

$H_A$ : not  $H_0$

Table 4.3 Result of Model Fit Statistics for Intercept only (empty) and Full Model

| Goodness of fit measure(for factor model) | Empty model | Full model |
|-------------------------------------------|-------------|------------|
| Log likelihood(LL)                        | -2921.5109  | -2194.5065 |
| AIC                                       | 5845.022    | 4427.013   |
| BIC                                       | 5852.35     | 4566.247   |

From Table 4.3, we can see that  $-2LL$  of the full model is  $-2(-2194.5065) = 4,389.013$  and  $-2LL_0$  of the intercept only model is  $-2(-2921.5109) = 5,843.0218$  giving the difference of the deviances is  $(5843.0218-4389.013) = 1454.0088$ . Since log likelihood ratio test is Chi-square distributed with 18 degrees of freedom,  $P(\chi^2(18) > 1454.0088) = 0.0000$ , we can reject the null hypothesis of no significant difference between the two models. We use AIC and BIC values to determine the better model. Table 4.3 shows that the full model is better than the empty model because AIC and BIC are smaller than for the empty model.

The overall goodness of fit of the model was evaluated by Hosmer-Lemeshow goodness of fit test with the null hypothesis:

$H_0$ : the model is a “good enough” fit to the data and we only reject the null hypothesis and decide it is “poor” fit if there are sufficiently strong grounds to do so (conventionally if  $p < 0.05$ ).

Table 4.4: Test of Significance of Hosmer-Lemeshow Goodness of Fit Statistics

| Hosmer-Lemeshow goodness of fit test |    |           |
|--------------------------------------|----|-----------|
| Chi-square                           | DF | Prob>chi2 |
| 9.29                                 | 8  | 0.3182    |

The p-value as shown in the table (0.3182) is greater than 0.05. Hence the model is a good fit to the data.

We use the Wald test to check the overall significance of the coefficients and individual coefficients in the model. The values of the Wald test statistic are given in the 4<sup>th</sup> column of Table 4.2 and are used to test the hypothesis that an individual coefficient is zero. The p-values are given in 5<sup>th</sup> column of Table 4.2.

Table 4.5 Wald test of all BETAs=0

| Chi square | DF | Prob>chi2 |
|------------|----|-----------|
| 721.11     | 17 | 0.0000    |

Since  $p=0.0000$  as shown in Table 4.5, the global null hypothesis:

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$  is rejected and we conclude that at least one coefficient is different from zero.



### 4.3.2 Interpretation of parameter estimates

For a dichotomous variable the parameter of interest is the odds ratio. An estimate of this parameter is obtained from the estimated logistic regression coefficient, regardless of how the variable is coded. This relationship between the logistic regression coefficient and the odds ratio provides the foundation for our interpretation of all logistic regression results.

When both the dependent and independent variables are both dichotomous, the odds ratio is the Probability that Y is 1 when X is 1 compared to the probability that Y is 1 when X is 0.

Logistic regression analysis of the contribution of socio-economic and demographic variables on under-five child mortality shows that husband/partner's education and occupation, the presence of TV in the household, mother's occupation, religion, place of residence, wealth index, marital status of mothers, sex of household head, age of mothers and sex of child had no significant effect on under-five mortality. In contrast, the length of breast feeding, number of household members, age of the mother, number of children ever born to the mother, mother's educational level, birth order number and region had significant effect on under-five child mortality.

Odds ratio is a measure of association; it is an estimate of the risk of an exposed group relative to control group or unexposed (reference) group. Odds ratio less than 1 indicates negative relationship and odds ratio greater than 1 indicates positive relationship and odds ratio=1 indicates no relationship.

Table 4.3 shows that the risk of dying before the age of 5 for a child breast fed for 0-6 months was nearly zero( $OR=0.0003$ ) relative to the risk for never breastfed under-five child controlling other variables in the model. Children breast fed for 7-12 months were 99.5% less likely to die before the age of 5 than never breast fed children. The risk of dying before the age 5 for an ever breast fed child was 94.4% less than a risk of dying for a never breast fed child controlling other variables in the model. Birth order number of 4-6 increased the exposure of under-five child mortality by about 72% ( odds ratio=1.718) as compared to birth order number of 1-3 and under-five child mortality increases by 82.7% (odds ratio=1.827) for birth order 7-9 in comparison with birth order number of 1-3. On the other hand, a child from families with 5-8 children was 51.9% more likely to die before the age of 5 compared to a child born to families with 1-4 ever born children controlling for other variables in the model. The exposure of under-five child mortality for families with at least 9 ever born children was about 3.35(odds ratio=3.352) times that for families with 1-4 ever born children to the mother.

Children in households of size 6-10 were 61% (odds ratio=0.39) less likely to die before the age of 5 compared to children in households of size 1-5. Children in households of size 11-15 were 75% less likely to die before the age of 5 compared to children in households of size 1-5. Children born to mothers with no education were more exposed to mortality than children born to mothers with higher education. That is, the risk of dying before the age of 5 of children born to uneducated mothers was about 6.5 (odds ratio=6.536) times the risk of dying of children born to mothers having higher education. Children born to mothers with primary education were about 6 (odds ratio=6.103) times more likely to die before the age of 5 than children born to mothers with higher

education. When we look at the regional differences in under-five child mortality, the odds ratio of Afar is 2.941 implying that under-five child mortality in Afar is about 3 times higher than that of Addis Ababa and the risk of under-five child mortality in Amhara and Oromia was about 2 (odds ratio of 2.273 and 2.153 respectively) times higher than that of Addis Ababa showing that Amhara and Oromia had nearly the same rate of under-five child mortality. Unexpectedly, under-five child mortality in Somali was not significantly different from that in Addis Ababa. This may be due to none coverage of all sample units as stated in EDHS 2011 final report. But children born in Benishangul-Gumuz had a risk of mortality about 3 (odds ratio=2.6) times higher than those of children born in Addis Ababa. Children in SNNP were about 3(odds ratio=2.826) times more likely to die before the age of 5 compared to children in Addis Ababa. This implies that Benishangul-Gumuz and SNNP had almost the same level of under-five child mortality compared to Addis Ababa. The exposure of under-five child mortality in Gambela was 2.4 (odds ratio=2.436) times higher than that for Addis Ababa. The likelihood of dying before the age of 5 for children born to mothers aged 15-19 years is about 5 (odds ratio=4.614) times more than the likelihood of dying for children born to mothers aged 45-49 years.

#### **4.3.3 Model diagnostics**

So far, we have discussed some summary statistics examined to check the goodness of fit of our model. Before concluding that the model is adequate, it is a common practice to examine other measures used to detect the presence of outliers and influential observations among others.

As can be seen in the Appendix values of the model diagnostic measures are less than unity. DFBETAs less than unity imply no specific impact of an observation on the coefficient of a particular predictor variable, Cook's distance less than unity showed that an observation had no overall impact on the estimated vector of regression coefficients  $\beta$ . A value of the leverage statistic less than one shows that no subject has a substantial large impact on the predicted values of the model (see the Appendix). Based on the above goodness of fit tests and diagnostic checking results, we can say that our model is adequate.

#### **4.4 Multilevel logistic regression analysis**

In this study two-level structures with mothers as first level units and regions as the second level units has been used. Mothers selected randomly from the same region are assumed to be more similar than mothers selected randomly from different regions. In this section three multilevel logistic regression models (namely empty model with random intercept, random intercept with fixed effects model and random coefficient with random intercept model) have been considered and model comparison, testing goodness of fit and interpretation of fixed effects in terms of odds ratio are presented.

For the proper application of multilevel analysis in general and multilevel logistic analysis in particular the first logical step is to test for heterogeneity of proportions between groups (in our case between regions). The chi-square test was applied to assess heterogeneity among regions. The test yields  $\chi^2 = 26.8208$  with d.f 10 ( $P=0.003$ ). Thus, there is evidence of heterogeneity among the regions.

#### 4.4.1 The Empty model with random intercept

The empty two-level model also called null two-level model for a dichotomous outcome variable refers to a population of groups (level-two unit i.e regions) and specifies the probability distribution for group-dependent probabilities,  $\pi_j$ . It is the model that incorporates only the grand mean and random intercept (regional effect) without covariate.

It is given by:  $\text{logit}(\pi_j) = \beta_0 + U_{0j}$ , where  $U_{0j} \sim \text{IID}(0, \sigma_u^2)$ .

The intercept  $\beta_0$  also known as the grand mean is shared by all regions while the random effect  $U_{0j}$  also known as level two residual is specific to region  $j$ . It shows how the mean of under-five children mortality in a particular region deviates from the grand mean.

$\sigma_u^2$  is the between regions variance. The random effect is not directly estimated but is summarized in terms of their estimated variances.

Table 4.6 shows the output of the estimates of fixed effects and random effects. From the table we can see that the estimate of the fixed part of the model is -2.58 with z-value of -40.95 and p-value of 0.000 which implies that the average log odds of under-five child mortality is significantly different from zero. The fixed part of the model is interpreted as the grand mean of log odds of under-five child mortality with the odds of  $\exp(-2.58)$

$= 0.076$  and the average probability of under-five child mortality  $\frac{\exp(-2.58)}{1 + \exp(-2.58)} = 0.0704$

which means the chance of under-five child mortality is 7.04% on average. The table also

contains the variance estimate of random effects at regional level,  $\sigma^2_u = 0.0236$  with confidence interval of [0.0037, 0.1523) which implies that the between region variance of under-five child mortality is 0.0236. At the bottom of the table there is the result of the hypothesis  $H_0: \sigma^2_u = 0$  is provided showing that there is no cross-regional variation in under-five children mortality. For this hypothesis, we see that the value of the test statistic is 3.11 with  $p = 0.0388$ . Therefore, the null hypothesis is rejected and there is evidence of heterogeneity or cross-regional variation in under- five children mortality.

Table 4.6 stata xtlogit output for random intercept only model

|                                                |             |           |           |                      |                         |            |
|------------------------------------------------|-------------|-----------|-----------|----------------------|-------------------------|------------|
| U5DIED                                         | Coef.       | Std. Err. | z         | P> z                 | [95% Conf. Interval]    |            |
| __cons                                         | -2.581583   | .0630373  | -40.95    | 0.000*               | [-2.705134              | -2.458033] |
| Random-effects Parameters                      |             | Estimate  | Std. Err. | [95% Conf. Interval] |                         |            |
| REGION: Identity                               |             |           |           |                      |                         |            |
|                                                | var(__cons) | .0235859  | .0224429  | [.0036534 .152266]   |                         |            |
| LR test vs. logistic regression: chibar2(01) = |             |           |           | 3.11                 | Prob>=chibar2 = 0.0388* |            |
| Model                                          | Obs         | LL value  |           | df                   | AIC                     | BIC        |
| m0                                             | 11250       | -2919.954 |           | 2                    | 5843.909                | 5858.565   |

We can now write the model for the  $j^{th}$  region as  $\text{logit}(\pi_j) = -2.58 + U_{0j}$ . From the model we can say that the average probability of under-five child mortality in the absence of covariates in region  $j$  is less than the average when  $U_{0j}$  is negative while it is higher than the average when  $U_{0j}$  is positive.

Table 4.7 shows the estimated values of  $U_{0j}$ . The results indicate that the probability of under-five child mortality is below the average (7%) in Tigray,Harari, Addis Ababa and Diredawa while it is higher than the average in the remaining regions. The worst situation has been in Affar and Benishangul-Gumuz.

Table 4.7 Output of the estimate of random effect for each region.

| REGION     | $\hat{u}_i$ |
|------------|-------------|
| Tigray     | -.0196429   |
| Affar      | .1494067    |
| Amhara     | .0240968    |
| Oromiya    | .0089426    |
| Somali     | .0470699    |
| Benishangu | .1953026    |
| SNNP       | .0471896    |
| Gambela    | .0122161    |
| Harari     | -.0248436   |
| Addis Abab | -.2194424   |
| Dire Dawa  | -.1949519   |

From Table 4.6 we can calculate intra-class correlation coefficient which is the measure of the correlation between two individuals who are in the same higher level unit (region). A low ICC indicates relatively small between region variations. From the table we have between regions variance of 0.0236 and level one variance of  $\pi^2/3 = 3.29$ . From equation (3.24) intra-class correlation coefficient is 0.00712 which is very small in fact.

#### 4.4.2 Random intercept model with fixed coefficients

We now extend our discussion by including level-one covariates in the aforementioned model. The model obtained in this way is called multilevel logistic model with random intercept and fixed slope.

In random intercept and fixed slope model, we allow the probability of under-five child mortality to vary across regions while mother's level covariates including the fixed intercept are fixed or constant across regions. Table 4.8 contains the maximum likelihood estimates of fixed slopes and regional effects of the model.

Table 4.8 stata xtmelogit output for random intercept with fixed slope model

| Variables with categories                                                  | Coef.    | Odds ratio | Std. Err of coef.    | Z      | p> Z   | [95%Conf.Interval for odds ratio] |        |
|----------------------------------------------------------------------------|----------|------------|----------------------|--------|--------|-----------------------------------|--------|
| Months of breast feeding                                                   |          |            |                      | 25.06  | 0.000* |                                   |        |
| 0-6 Months                                                                 | -6.809   | .0004      | 1.007                | -7.76  | 0.000* | .000                              | .003   |
| 7-12 months                                                                | -5.234   | .005       | .354                 | -14.79 | 0.000* | .003                              | .011   |
| 13-18 months                                                               | -6.694   | .0012      | .717                 | -9.34  | 0.000* | .0003                             | .005   |
| 19-59 months                                                               | -6.782   | .0011      | .589                 | -11.52 | 0.000* | .00004                            | .0036  |
| Ever breast fed                                                            | -2.83    | .059       | .116                 | -24.38 | 0.000* | .047                              | .074   |
| Never breast fed(ref)                                                      |          |            |                      |        |        |                                   |        |
| Birth order number                                                         |          |            |                      | 1.25   | 0.02*  |                                   |        |
| 1-3(ref)                                                                   |          |            |                      |        |        |                                   |        |
| 4-6                                                                        | .548     | 1.730      | .187                 | 2.94   | 0.003* | 1.200                             | 2.495  |
| 7-9                                                                        | .629     | 1.876      | .266                 | 2.36   | 0.018* | 1.12                              | 3.162  |
| 10 and above                                                               | -.0412   | .960       | .407                 | -0.10  | 0.929  | .432                              | 2.129  |
| Age of mother at first birth                                               |          |            |                      | 1.56   | 0.116  |                                   |        |
| 9-14                                                                       | -.800    | .449       | .402                 | -1.99  | 0.047* | .204                              | .989   |
| 15-19                                                                      | -.689    | .502       | .368                 | -1.87  | 0.061  | .244                              | 1.032  |
| 20-24                                                                      | -.590    | .554       | .356                 | -1.65  | 0.098  | .276                              | 1.115  |
| 25-29                                                                      | -.528    | .590       | .383                 | -1.38  | 0.168  | .279                              | 1.249  |
| 30 and above(ref)                                                          |          |            |                      |        |        |                                   |        |
| Total children ever born                                                   |          |            |                      | 4.09   | 0.000* |                                   |        |
| 1-4(ref)                                                                   |          |            |                      |        |        |                                   |        |
| 5-8                                                                        | .421     | 1.523      | .183                 | 2.30   | 0.022* | 1.064                             | 2.180  |
| 9 and above                                                                | 1.209    | 3.351      | .284                 | 4.26   | 0.000* | 1.921                             | 5.844  |
| Number of house hold members                                               |          |            |                      | -8.63  | 0.000* |                                   |        |
| 1-5(ref)                                                                   |          |            |                      |        |        |                                   |        |
| 6-10                                                                       | -.944    | .389       | .106                 | -8.87  | 0.000* | .315                              | .479   |
| 11-15                                                                      | -1.429   | .239       | .275                 | -5.2   | 0.000* | .140                              | .411   |
| 16 and above                                                               | -.386    | .680       | .514                 | -0.75  | 0.453  | .248                              | 1.864  |
| Mothers educational level                                                  |          |            |                      | -3.34  | 0.001* |                                   |        |
| No education                                                               | 2.007    | 7.439      | .625                 | 2.21   | 0.001* | 2.186                             | 25.320 |
| Primary                                                                    | 1.868    | 6.467      | .627                 | 2.98   | 0.003* | 1.894                             | 22.141 |
| Secondary                                                                  | 1.355    | 3.876      | .678                 | 2.00   | 0.046* | 1.027                             | 14.626 |
| Higher(ref)                                                                |          |            |                      |        |        |                                   |        |
| Age of mother                                                              |          |            |                      | -4.71  | 0.000* |                                   |        |
| 15-19                                                                      | 1.441    | 4.225      | .396                 | 3.64   | 0.000* | 1.943                             | 9.184  |
| 20-24                                                                      | .777     | 2.175      | .350                 | 2.22   | 0.026* | 1.096                             | 4.318  |
| 25-29                                                                      | .362     | 1.436      | .322                 | 1.12   | 0.262  | .763                              | 2.700  |
| 30-34                                                                      | .001     | 1.001      | .303                 | 0.000  | 0.996  | .552                              | 1.815  |
| 35-39                                                                      | .171     | 1.187      | .282                 | 0.61   | 0.543  | .683                              | 2.062  |
| 40-44                                                                      | -.246    | .782       | .296                 | -0.83  | 0.407  | .437                              | 1.398  |
| 45-49(ref)                                                                 |          |            |                      |        |        |                                   |        |
| Constant                                                                   | -4.045   |            | .728                 | -5.55  | 0.000* |                                   |        |
| Random-effects Parameters                                                  | Estimate | Std. Err.  | [95% Conf. Interval] |        |        |                                   |        |
| REGION: Identity                                                           |          |            |                      |        |        |                                   |        |
| var(_cons)                                                                 | .034     | .026       | .008                 | .148   |        |                                   |        |
| LR test vs. logistic regression: chibar2(01) = 6.84 Prob>=chibar2 = 0.0045 |          |            |                      |        |        |                                   |        |

“\*” indicates significance for  $p < 0.05$



Table 4.9 model selection criteria and likelihood ratio test

|                                                        | Empty model with random intercept             | Random intercept with fixed effects model |
|--------------------------------------------------------|-----------------------------------------------|-------------------------------------------|
| Log likely hood(LL)                                    | -2919.954                                     | -2122.936                                 |
| -2LL=deviance                                          | 5839.908                                      | 4245.872                                  |
| AIC value                                              | 5843.909                                      | 4301.872                                  |
| BIC value                                              | 5858.565                                      | 4507.059                                  |
| Likelihood-ratio test<br>(Assumption: m2 nested in m1) | LR chi2(26) = 1594.04<br>Prob > chi2 = 0.0000 |                                           |

The table contains estimates of the fixed slopes and associated odds ratios. Values of the Wald test statistic used for testing for the significance of individual predictors are given in the 5<sup>th</sup> column with the corresponding p-values in the 6<sup>th</sup> column. The Wald test of overall goodness of fit gives Wald chi2 (26) =880.94 with p =0.000 where 26 is the degrees of freedom. This indicates that all explanatory variables jointly are significant. From the table we see that the inclusion of level one covariates increased regional variations from 0.0236 (level-two variance without covariates) to 0.034 which may not be in fact expected. The likelihood ratio test in Table 4.9 shows that there is a significant difference between the random intercept only model and random intercept with fixed slope model.

The deviance of the random intercept model, 5839.908 is reduced to 4245.872 when we include covariates for the same random intercept which implies that the random intercept with fixed slope model is better than the random intercept only model. The BIC and AIC values in Table 4.9 also ensure this as the smaller the values of AIC and BIC the better the model is.

Moreover, the values of chi2 (1) =6.84 and p=0.0045 (see Table 4.8) lead to the rejection of the null hypothesis that the random effect is zero as in the assumption of ordinary

logistic regression. From this we can conclude that the random effect at regional level is significantly different from zero.

From Table 4.8 we see that all categories of months of breastfeeding (0-6 months, 7-12 months, 13-18 months, 19-59 months and ever breastfed but have discontinued at the time of the survey) are significant for U5CM as compared to the reference (never breastfed) category. In addition birth order of 4-6, 7-9, number of children ever born (5-8 and 9 and above), number of household members (6-10 and 11-15) are significant factors for U5CM as compared to their reference categories.

Mothers' educational level (no education, primary education and secondary education) and age of mother (15-19 years and 20-24) significantly affected U5CM compared to the reference categories.

A slight difference in the values of the odds ratio corresponding to the two models was observed. While the two categories of age of mother at first birth (9-14 and 15-19 years) were significant in the ordinary (single level) logistic regression, only one category (9-14 years) was significant in the random intercept model for the same reference category value. Similarly, all levels of education for random intercept with fixed effects model were significant compared to the reference level while secondary educational level in the ordinary logistic regression was not significant as compared to the reference level.

The most appropriate way of interpreting hierarchical logit model is odds ratio. There was almost no risk of dying before age 5 for a child born to mothers breastfed their children for 0-6 months relative to child born to mothers never breastfed their children. Children breastfed for 7-12 months were 99.5% less likely to die before the age of 5

(OR=0.005) as compared to children never breastfed controlling for the other variables in the model. Similarly the risk of dying before age 5 for children born to mothers who ever breastfed their children but have discontinued breast feeding at the time of the survey was 94.1% less than the risk of dying for children born to mothers who never breastfed their children controlling other variables in the model. The chance of under-five child mortality for birth order of 4-6 and 7-9 was 73% (OR=1.73) and 87.6% (OR=1.876) respectively higher than the chance of U5M for birth order of 1-3 when other variables are kept constant. Children born from mothers who gave 5-8 births and 9 or more children were 1.523 and 3.351 times more likely to die before the age of 5, respectively compared to children born to mothers with a total of 1-3 ever born children for constant effect of other variables. The risk of under-five child mortality for families of size 6-10 and 11-15 was 38.9% (OR=0.389) and 23.9% (OR=0.239) times respectively the risk of U5CM for households of size 1-5. On the other hand, the risk of experiencing under – five child mortality for mothers with no education was 7.439 times higher; about 6.5 (OR=6.467) times higher for mothers with primary education; about 3.9 (OR=3.876) times higher for mothers with secondary education than the risk of experiencing U5CM for mothers with higher education when the effect of other variables is controlled. Likewise, children born to 15-19 and 20-24 years old mothers were about 4.2 (OR=4.225) and 2.2 (OR=2.175) times more likely to die as compared to the reference category keeping the effect of other variables constant.

#### **4.4.3 Random intercept with random coefficients model**

Multilevel logistic regression can allow the coefficient of level-one covariates to vary across regions instead of keeping them fixed across regions. Now we are going to see the

effect of mothers' level covariates by allowing them to vary randomly across regions. This model contains fixed effects and random effects. The fixed effects are analogous to standard logistic regression coefficients and are estimated directly. The random effects are not directly estimated but are summarized in terms of their estimated variances and covariances. The random effects can take random intercepts (regional effects) and random coefficients (level-one covariates effect). In this section we investigate whether level-one covariates have random effects across regions or they have the same effects across regions. Estimates of this model showed that the variance of random slopes of all included variables is zero except for length of months of breast feeding and age of mother. This indicates that only the effects of length of breast feeding and age of mother on under-five child mortality varied across regions whereas the effect of other covariates for under-five child mortality remained fixed across regions. The results of the random slope estimates are given in Table 4.10 below.

Table 4.10 stata extmelogit outputs of random slopes model.

| U5DIED                       | Coef.  | Std.E<br>rr. | Z      | P>z    | Odds<br>ratio | 95% CI.For OR |       |
|------------------------------|--------|--------------|--------|--------|---------------|---------------|-------|
| Months of breast feeding     |        |              | 25.02  | 0.000* |               |               |       |
| 0-6 months                   | -8.188 | 1.145        | -7.15  | 0.000* | .0003         | .000          | .0026 |
| 7-12 months                  | -5.367 | .507         | -10.58 | 0.000* | .005          | .002          | .013  |
| 13-18 months                 | -6.782 | .765         | -8.87  | 0.000* | .001          | .0002         | .005  |
| 19-59 months                 | -6.805 | .613         | -11.10 | 0.000* | .001          | .0003         | .003  |
| Ever breast fed              | -2.861 | 141          | -20.34 | .000*  | .057          | .043          | .075  |
| Never breast fed(ref)        |        |              |        |        |               |               |       |
| Birth order number           |        |              | 1.27   | 0.024* |               |               |       |
| 1-3(ref)                     |        |              |        |        |               |               |       |
| 4-6                          | .559   | .187         | 2.99   | 0.003* | 1.748         | 1.212         | 2.523 |
| 7-9                          | .638   | .267         | 2.39   | 0.017* | 1.893         | 1.121         | 3.198 |
| 10 and above                 | -.01   | .407         | -0.02  | 0.981  | .990          | .446          | 2.200 |
| Age of mother at first birth |        |              | 1.52   | 0.113  |               |               |       |
| 9-14                         | -.820  | .403         | -2.01  | 0.045* | .445          | .202          | .980  |
| 15-19                        | -.695  | .368         | -1.89  | 0.059  | .499          | .243          | 1.027 |

|                              |          |           |                     |        |       |       |        |
|------------------------------|----------|-----------|---------------------|--------|-------|-------|--------|
| 20-24                        | -.594    | .356      | -1.67               | 0.095  | .552  | .274  | 1.110  |
| 25-29                        | -.523    | .383      | -1.37               | 0.172  | .593  | .280  | 1.256  |
| 30 and above(ref)            |          |           |                     |        |       |       |        |
| Total children ever born     |          |           | 4.06                | 0.000* |       |       |        |
| 1-4(ref)                     |          |           |                     |        |       |       |        |
| 5-8                          | .417     | .184      | 2.27                | 0.023* | 1.518 | 1.059 | 2.174  |
| 9 and above                  | 1.214    | .285      | 4.26                | 0.000* | 3.366 | 1.926 | 5.882  |
| Number of house hold members |          |           | -8.63               | 0.000* |       |       |        |
| 1-5(ref)                     |          |           |                     |        |       |       |        |
| 6-10                         | -.947    | .107      | -8.87               | 0.000* | .388  | .314  | .478   |
| 11-15                        | -1.413   | .273      | -5.17               | 0.000* | .243  | .142  | .4159  |
| 16 and above                 | -.417    | .511      | -0.82               | 0.415  | .659  | .242  | 1.795  |
| Mothers educational level    |          |           | -3.34               | 0.001* |       |       |        |
| No education                 | 2.035    | .631      | 3.22                | 0.001* | 7.653 | 2.222 | 26.362 |
| Primary                      | 1.890    | .634      | 2.98                | 0.003* | 6.623 | 1.913 | 22.926 |
| Secondary                    | 1.355    | .684      | 1.98                | 0.047* | 3.878 | 1.015 | 14.809 |
| Higher(ref)                  |          |           |                     |        |       |       |        |
| Age of mother                |          |           | -3.79               | 0.000* |       |       |        |
| 15-19                        | 1.414    | .432      | 3.27                | 0.001* | 4.113 | 1.762 | 9.599  |
| 20-24                        | .779     | .380      | 2.05                | 0.040* | 2.179 | 1.036 | 4.584  |
| 25-29                        | .376     | .344      | 1.09                | 0.275  | 1.456 | .742  | 2.856  |
| 30-34                        | .0132    | .318      | 0.04                | 0.967  | 1.013 | .543  | 1.891  |
| 35-39                        | .172     | .290      | 0.59                | 0.554  | 1.187 | .672  | 2.098  |
| 40-44                        | -.220    | .300      | -0.73               | 0.464  | .803  | .446  | 1.446  |
| 45-49(ref)                   |          |           |                     |        |       |       |        |
| Constant                     | -4.073   | .738      | -5.52               | 0.000* |       |       |        |
| Random-effects Parameters    | Estimate | Std. Err. | [95% Conf Interval] |        |       |       |        |
| REGION: Unstructured         |          |           |                     |        |       |       |        |
| var(AGEM)                    | .007     | .007      | .001                | .051   |       |       |        |
| var(MBFEED)                  | .046     | .046      | .006                | .330   |       |       |        |
| var(_cons)                   | 1.773    | 1.568     | .313                | 10.041 |       |       |        |
| cov(AGEM,MBFEED)             | .016     | .013      | -.009               | .0406  |       |       |        |
| cov(AGEM, _cons)             | -.108    | .083      | -.271               | .056   |       |       |        |
| cov(MBFEED, _cons)           | -.278    | .266      | -.800               | .243   |       |       |        |

**LR test vs. logistic regression:  $\chi^2(6) = 14.29$  Prob >  $\chi^2 = 0.0265$**

Note: ref indicates reference and ‘\*’ indicates significance at 5% significance level

Wald  $\chi^2(26) = 491.75$  with p-value=0.000 indicates that at least one population parameter is significantly different from zero. Moreover, the likelihood ratio (LR) test

shown at the foot of the table provides evidence that the logistic regression model with no intercept is rejected by the data. This means that the variance-covariance of random effects of the population is significantly different from zero.

The estimate of the fixed intercept is -4.073 and the log-odds of the probability of under-five child mortality when all level one covariates are zero in region  $j$  is given by  $-4.073+\hat{u}_j$  where  $\hat{u}_j$  is a random intercept with variance of 1.773(indicated in the table as var(cons)) which is the between-regions variance and standard error 1.568. In the absence of level-one covariates, the status of each region on under-five child mortality as compared to the average child mortality measured with log odds depends on the sign of the random intercept,  $\hat{u}_j$ . When  $\hat{u}_j$  is positive the log odds of under-five child mortality is higher than the average and when  $\hat{u}_j$  is negative the log odds of under-five child mortality is less than the average. The individual region slopes of months of breast feeding (MBFEED) and age of mother (AGEM) vary with variance 0.046 and 0.007, respectively.

The correlation matrix contains the estimated correlation between random intercepts and the random slopes. Positive correlation between intercepts and slopes implies that regions with higher intercepts tend to have on average higher slopes. The negative sign for the correlation between intercepts and slopes implies that regions with higher intercepts tend to have on average lower slopes on the corresponding predictors (see Table 4.11).

Table 4.11 Random-effects correlation matrix for level REGION

|        | AGEM  | MBFEED | Cons |
|--------|-------|--------|------|
| AGEM   | 1     |        |      |
| MBFEED | .864  | 1      |      |
| Cons   | -.958 | -.972  | 1    |

In order to see whether the inclusion of level one covariates with age of mother and duration of breastfeeding varying across regions significantly improved the random intercept only model, log likelihood ratio test, AIC and BIC values have been used. The likelihood ratio test results are  $LR\ chi^2(31) = 1601.49$  (which is the difference between the deviance of random intercept only model and random intercept with random slopes model ;  $5839.908 - 4238.416$ ) and  $p = 0.000$  implying that there is a significant difference between the two nested models. The AIC( 4304.417)and BIC(4546.245) values for the random intercept with random slope model are less than the corresponding AIC(5843.909) and BIC(5858.565) values for the random intercept only model indicating that the random intercept with random slope model is better than random intercept only model (see table 4.12).

Table 4.12 results of model selection criteria and log likelihood ratio test

| Model selection criteria | Random intercept only model                        | Random intercept and random slope model |
|--------------------------|----------------------------------------------------|-----------------------------------------|
| Log likelihood           | -2919.954                                          | -2119.208                               |
| Deviance                 | 5839.908                                           | 4238.416                                |
| AIC                      | 5843.909                                           | 4304.417                                |
| BIC                      | 5858.565                                           | 4546.245                                |
| Likelihood-ratio test    | $LR\ chi^2(31) = 1601.49$ $Prob > \chi^2 = 0.0000$ |                                         |

The likelihood ratio test of random intercept model versus random intercept with random coefficient model yields  $LR\ chi^2(5) = 7.45$  which is the difference between the two deviances and  $p = 0.1890$  (where 5 is the degrees of freedom obtained by taking the difference of the number of parameters in the two models) which implies that the null hypothesis of no difference between the two models is not rejected. Both the AIC



(4301.872) and BIC(4507.059) values of the random intercept model in Table 4.9 are less than the AIC(4304.417) and BIC(4546.245) values of the random intercept with random slopes model as shown in Table 4.12 indicating that the random intercept or covariance model is preferred though the difference between the two models is not significant.

## **CHAPTER FIVE**

### **5. Discussion, conclusion and recommendation**

#### **5.1 Discussion and conclusion**

This study intended to identify the most important socio-economic and demographic determinants of under-five child mortality based on the 2011 EDHS data for the last five years before the survey. The descriptive analysis results indicated that under-five child mortality varied among the regions. The highest U5C mortality was observed in Benishangul-Gumuz followed by Afar and the least under 5 child mortality was observed in Addis Ababa followed by Diredawa. This disparity could be due to unequal distribution of infrastructures and different socio-economic and demographic characteristics exercised in each region. Under-five child mortality showed an inverse relationship with education of mothers. As the level of education of mothers increased, child mortality rate decreased. Under-five child mortality showed unexpected U-shaped result corresponding to the number of household members in that the highest mortality was recorded in households of size 1-5 and 16 and more respectively. The lowest under-5 child mortality was observed in the households of size 11-15 and 6-10 respectively.



Whether head of the household is male or female did not show any effect in under 5 child mortality while wealth index of the household had an inverse effect on under-five child mortality in that households having higher living standards had less under-five child mortality. Age of mother at first birth showed the expected result that highest under-5 child mortality occurred for mothers aged 30 and above, 9-14 and 15-19 years and the least under 5 child mortality were observed for mothers in the age group 25-29 years which may be consistent with biological reasoning. A surprising result was that under-five child mortality was lowest for mothers no longer living with their husband/partner. Birth order number showed a U-shaped relationship with under-five child mortality which is consistent with the result obtained by Desta Mekonen(2011) in that under-five child mortality is high for lower birth order and higher birth order.

Professional/technical worker mothers experienced the least proportion of under-five child mortality as compared to nonprofessional/technical mothers. This could be due to higher level of education of professional/technical worker mothers. An interesting descriptive result was obtained in months of breastfeeding. Child mortality was lowest for those breastfed for 6 months and mortality increased slightly for increased months of breast feeding. The reason could be exclusive breastfeeding without complimentary foods after 6 months. Conversely, maximum under-5 mortality was observed for those who never breastfed. Child mortality was high for young mothers (below 20 years) and old mothers (older than 40 years) and our descriptive analysis ascertained this. Consistent with the report by UNICEF (2012), child mortality was higher in rural place of residence than urban place of residence.

The results of the cross tabulation analysis showed that place of residence significantly affected under-five child mortality similar to the findings of Kummar and Gemechis (2010). This is due to the fact that children born in urban areas have access to health facilities and are in a better position to get proper nourishment and mothers are more exposed to mass media to learn how to care for their children. Similarly, level of mother's education and under-five child mortality had significant relationship which is in agreement with the findings by Mondal (2009). This is because higher level of education of women greatly changes family relationships positively and brings profound effects on child care. Mother's education reinforces her ability to make use of government and private health care resources and it may increase the autonomy necessary to advocate for her child in the household and the outside world (Caldwell 1989). On the contrary husband /partner's educational level had no significant role in under-five child mortality which is in line with the findings of Mturi and Curtis (1995). This could be due to the increasing independent decision making power of mothers and fathers. Similarly, the availability of television in the household, religion of mothers and sex of household head, age of mother at first birth and birth order number were not significantly associated with under-five child mortality.

The number of household members and wealth index of the household were significantly affecting under-five child mortality. Likewise, consistent with the results obtained by Tawiah (1979), both husband and mother's occupation showed a significant effect on under-five child mortality. This may be reasonable as the type of occupation determines the standard of living of the household and wellbeing of the parents. On the other side,

mother's occupation determines the length of time she spends with her child and provides care to her children

Duration of breast feeding had an effect on under-five child mortality. Exclusive breast feeding for 6 months and with complimentary foods for the next months reduces mortality before age 5 as agreed by health professionals. Furthermore, the result that age of mothers played a significant role in under-five child mortality is consistent with the study by Mondal(2009).Additionally, sex of child and under-five child mortality were significantly associated.

The logistic regression analysis confirmed that breastfeeding was an important factor in reducing under-five child mortality. This is consistent with the conclusion given by Mondal (2009), Kwabena (2011), Tawiah (1979) and Desta Mekonnen (2011). Women living in developing countries have experience of prolonged breast feeding in alternative to other nutrition. Other factors being constant, birth order was a significant factor for under-five child mortality. This result is consistent with the results obtained by Desta (2011). In this study it was shown that birth order 4-6 and 7-9 significantly affected under-five child mortality compared to the 1-3 birth order number. Total number of children ever born to the mother had a significant effect on under-five child mortality which is in line with the conclusion given by Kwabena (2011). Under-five child mortality increased by about 1.5 and 3.4 times in households with 5-9 and at least 9 children ever born compared to households with 1-4 children ever born. As the number of children ever born to a mother increased, the more probable that under-five child mortality will occur because of insufficient supply of food and inadequate care the child received from its mother. Other factors being constant, number of household members significantly

affected under-five child mortality. Households of size 6-10 and 11-15 were affecting under-five child mortality compared to households of size 1-5 Desta Mekonnen (2011), using EDHS 2005 data reported similar finding.

Education plays a significant role in reducing under-five child mortality, particularly maternal education has a greater contribution towards the reduction of under-five child mortality. In line with this, Caldwell (1989) found that infant and child mortality in Nigeria was highly associated with mother's education. In India, Avind Pandey, et al (1998) found a significance relationship of mother's education and infant and child mortality. Kwabena (2011) in South Africa and Kumar and Gemechis (2010) in Ethiopia found similar results that mother's education and under-five child mortality were significantly associated. This study found that children born to mothers with no education were highly exposed to mortality than children born to mothers with higher education. However, father's level of education, which can play a crucial role in reducing under-five child mortality, was not significantly affecting under-five child mortality.

It is expected that regional variations exist in under-five child mortality due to unequal distribution of infrastructures and health care facilities and different socio-economic and cultural experiences. This study showed that disparities of under-five mortality prevailed among regions. A similar result was found in Kenya by Mustafa and Odimegwu (2008). Households in Afar, Amhara, Oromia, Benishangul-Gumuz, SNNP and Gambela experienced higher under-five child mortality than households in Addis Ababa. But unexpected result was obtained in Somalia region in that it was not significantly different from Addis Ababa

Children born to very young mothers and older mothers were likely to have higher risks of mortality. Very young mothers may experience difficult pregnancies and deliveries because of their physical immaturity. They are also likely to have limited knowledge and confidence in caring for infants and young children. Old mothers may also experience age-related problems during pregnancy and delivery. Thus we expect a U-shaped relationship between mother's age at childbirth and infant and child mortality (Arvind Pandey 1998). In line with this, age of mothers was significantly affecting under-five mortality.

Place of residence, possession of TV in the household, religion, sex of household head, wealth index, marital status of mother, both mothers and father's occupation and sex of child were not found significant. Similar significant effects were observed in the multilevel analysis except the fact that all educational levels of mothers had no positive effect on under-five child mortality as compared to higher educational level.

## **5.2 Conclusion**

Months of breast feeding was a predominant factor in reducing under-five child mortality and birth order number of 4-6 increased the exposure of under-five child mortality by about 72% compared to birth order number of 1-3. Under-five child mortality increased by 82.7% for birth order 7-9 in comparison with birth order number of 1-3. A child from families with 5-8 children was 51.9% more likely to die before the age of 5 compared to a child born to families with 1-4 ever born children and under-five child mortality for families with at least 9 ever born children was about 3.35 times that for families with 1-4 ever born children.

Mothers' educational level was also an important factor affecting under-five child mortality. That is, as the level of mothers' education increased, the level of mortality decreased. Children born to mothers aged 15-19 years were about 5 times more likely to die before the age of 5 than children born to mothers aged 45-49. Children born to mothers aged 20-24 were about 2.3 times more likely to die before the age of 5 than children born to mothers aged 45-49. The significant effect of number of household members was unexpected and further examination is recommended. The risk of dying before the age of 5 in Afar, Amhara, Oromia, Benishangul-Gumuz, SNNP and Gambela were significantly higher than that for Addis Ababa. The multilevel analysis region revealed that there was unobserved heterogeneity across regions and attention should be given to regions while formulating under-five child survival policies.

### **5.3 Recommendations**

Based on our findings, the following are recommended in order to reduce under-five child mortality.

- Breast feeding of new born babies for at least 6 months should be increasingly advocated.
- Exclusive breastfeeding practices for children aged 0-6 months should be promoted.
- Investments on infrastructures and health care facilities should be based on needs by each region.
- Efforts should be made to educate women to at least to secondary level.
- Mother and child health care services should be strengthened

- There should be a good supply of birth control methods to mothers for usage so that they give birth to small (not more than 4) number of children.

## References

- Agresti, A. (2007). An Introduction to Categorical Data Analysis. John Wiley And Sons,
- Arvind Pandey (1998). National Family Health Survey Subject Reports, Number 11
- Bawah. A. (2001). Living standards, Household size and Childhood Survival in Africa:
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed Models. *J. Am. Statist. Assoc.*, **88**, 9–25. *c Springer 2008*
- Caldwell JC (1989). Education as a factor in Mortality Decline: An Examination of Nigeria data. *Population Studies*, **33(2)**: 395-413.
- Central Statistical Authority and ORC Macro (2006). Demographic and Health Survey 2005, Ethiopia: Central Statistical Authority and MEASURE DHS.
- Cleland, J.G. and J.K. van Ginneken. (1988). maternal education and child survival in developing countries: the search for pathways of influence. *Social Science and Medicine* **27**,12:1357-1368
- Commenges, D. and Jacqmin, H. (1994). The intra-class correlation coefficient Distribution frees definition and test. *Biometrics*, **50**, 517-526.
- Cook (2009). History of the Social Determinants of Health: Global Histories, Contemporary Debates. Orient Blackswan.

- Cook, R. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, New York.
- Copas, J.B. (1988). Binary Regression Models for Contaminated Data. *Journal of Royal*
- DeLong, E.R., Peterson, E.D., DeLong, D.M., Muhlbaier, L.H., Hackett, S. and Mark, D.B. (1997). Comparing Risk-Adjustment Methods for Provider Profiling. *Statistics in Medicine*, **16**, 2645-2664.
- Desta Mekonnen (2011). Infant and Child Mortality in Ethiopia: The role of Socioeconomic, Demographic and Biological factors In the previous five years period of 2000 and 2005 developing countries. *Population and Development Review*; 10 (supplement).
- Ethiopia Demographic and Health Survey (2011). Central Statistical Agency Addis Ababa, Ethiopia ICF International Calverton, Maryland, USA.
- Ethiopia Ministry of Health. Health Sector Development Programme IV, 2010/11 – 2014/15.
- Federal Ministry of Health (2005). Essential Health Service Package for Ethiopia. Addis Ababa.
- Forum the Environment (2010). Ethiopian Environment Review; Addis Ababa, Ethiopia.
- Gelfand, A. E., Hills, S.E., Racine-Poon, A. and Smith A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**, 972–985.
- Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data. *Biometrika*, **78**, 45–51.
- Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *J. Roy. Statist. Soc. A*, **159**, 505–513.
- Goldstein, H., Multilevel Statistical Models. 3rd edition (2003) London: Arnold; New York: Oxford University Press Inc.
- Government of Ethiopia (2007). Report on Progress in Implementing the World Fit for Children Plan of Action in Ethiopia.
- Guang G. and Hongxin Zhao (2000). Multilevel modeling for binary data, *Annu. Rev. sociol.* 2000. **26**:441-62.



- Hill, K. (1985). The Pace of Mortality Decline since 1950 in Da Vanzo, Habicht, Hill and Preston, Quantitative Studies of Mortality Decline in the Developing World (World Bank Staff Working Papers Number 683, Population and Development Series Number 8). The World Bank, Washington, D.C. 1985.
- Hobcraft J.N. et al. (1984). Socio- economic factors in child mortality, a cross-national comparison, *Popul. Stud.*, **38**, 193-223.
- Hosmer, D. and Lemeshow, S. (2000, 2nd Ed). *Applied Logistic Regression*. John Wiley and Sons Inc., New York.  
Inc, New York
- Joop J. Hox(2010). *Multilevel Analysis Techniques and Applications*. Second Edition ,Utrecht University, The Netherlands
- Joshua Kembo & Jeroen K. Van Ginneken(2009). Determinants of infant and child mortality in Zimbabwe.  
▪ *Journal of American Statistical Association*, 92:162–170
- Kabir M, Chowdhury RI (1993). Infant and Child mortality levels and trends in Bangladesh: secondary analysis of the Bangladesh Fertility Survey, 1989. Dhaka: Bangladesh: National Institute of Population Research and Training (NIPORT).
- Khan, H. and Shaw, E. (2011). Multilevel logistic regression analysis applied to binary contraceptive prevalence data, *University of Warwick*.
- Kombo, J. and Ginneken, V. (2009). Determinants of infant and child mortality in Zimbabwe: Result of multivariate hazard analysis.
- Kumar, P. and Gemechis. (2010). Infant and child mortality in Ethiopia: As statistical analysis approach.
- Kwabena A. Kyei (2011). Socio – Economic Factors Affecting Under Five Mortality in South Africa – An Investigative Study
- Kyei, K. A. (1999). The mode of feeding that promotes higher survival among children. In UAPS. *The African Population in the 21st century*. Dakar: UAPS, (1) 193-212.
- Laird, N. M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, **65**, 581–590.
- McCulloch C. E., (1997) Maximum likelihood algorithms for generalized linear mixed models.

- McCullough, P. and Nelder, J.A. (1989). Generalized Linear Models. 2nd Edn. Chapman and Hall, London.
- MEOH (2010). HSDP IV 2010/2011-2014/2015
- Mondaal (2009). Factors Influencing Infant and Child Mortality: A Case Study of Rajshahi District, Bangladesh
- Mosley, W.H. and Chen, L.C. (1984). An analytical framework for the study of child survival in developing countries. *Population and Development Review* 10: 25-45.
- Mturi, A. J. and Curtis, L. S. (1995). The determinants of infant and child mortality in Tanzania. *Health policy plan* 10:384-94.
- Mustafa, E. and Odimegwu, C. (2008). Socioeconomic determinants of infant mortality in Kenya: Analysis of Kenya DHS 2003.
- Newsom (2011). Significance Testing in Multilevel Regression
- Preston, S.A. (1975). The Changing Relation between Mortality and Level of Economic Development *Population Studies*. 29(2):2231-2248.
- Preston, S.A. (1985). Mortality and Development Revisited in Da Vanzo, Habicht, Hill and Preston, *Quantitative Studies of Mortality Decline in the Developing World* (World Bank Staff Working Papers Number 683, Population and Development Series Number 8). The World Bank, Washington, D.C. 1985.
- Ridder G. and I. Tunalı (1999). Stratified partial likelihood estimation, *Journal of Econometrics*, 92: 193-232.
- Rio+20(2012): The United Nations Conference on Sustainable Development.
- Safeer Ullah (2010). Re-examining Socio-economic & Demographic determinants of High Infant Mortality in Pakistan, Analysis of PIHS 1998-99.
- Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling*. London/ Thousand Oaks/ New Delhi: Sage Publications  
*Statistical Association, B* 50 (2): 225-265.
- Tawiah, E. O. (1979). Some demographic and social differentials in Infant and early Childhood Mortality in Ghana. In *Population Dynamics, Fertility and Childhood Mortality in Africa*. Proceedings of the Experts Group Meeting on Fertility and Mortality levels and trends in Africa and their implications. Addis Ababa: UNECA
- Trussell J. and C. Hammerslough (1983). A hazards-model analysis of the covariates of infant and child mortality in Sri Lanka, *Demography*, 20: 1-26.

- UNICEF (2012). Levels and trends in child mortality reports 2012: United Nations Inter-agency Group for Child Mortality Estimation.
- UNICEF (2012). Progress report, division of policy and strategy; renewing the promise- in every country, for every child.
- United Nations Country Team, March (2011). Ethiopia United Nations Development Assistance Framework, 2012 to 2015.
- World Bank (2004). Ethiopia: A Country status report on health and poverty.
- World Bank. (1993). Investing in Health. Oxford University Press: New York, NY.
- World Health Statistics (2012). World health organization.

## Appendices

### Appendix A: results of cross-tabulation analysis

tab PR U5DIED,r chi2

| Type of place of residence | Under-five Child died |     | total  |
|----------------------------|-----------------------|-----|--------|
|                            | No                    | Yes |        |
| Urban                      | 1,781                 | 108 | 1,889  |
| rural                      | 8,664                 | 697 | 9,361  |
| Total                      | 10,445                | 805 | 11,250 |

Pearson chi2(1) = 7.0685 Pr = 0.008

tab MEDU U5DIED,chi2

| Mothers' educational leve | Under-five Child died |     | Total |
|---------------------------|-----------------------|-----|-------|
|                           | no                    | yes |       |
| No education              | 7288                  | 589 | 7877  |
| primary                   | 2618                  | 196 | 2814  |
| Secondary                 | 354                   | 17  | 371   |
| Higher                    | 185                   | 3   | 188   |
| total                     | 10445                 | 805 | 11250 |

Pearson chi2(3) = 13.8276 Pr = 0.003

tab HTV U5DIED, chi2

| House hold has TV     | Under-five Child died |     | total |
|-----------------------|-----------------------|-----|-------|
|                       | no                    | yes |       |
| No                    | 9169                  | 725 | 9894  |
| yes                   | 1034                  | 63  | 1097  |
| Not a dejure resident | 242                   | 17  | 259   |
| total                 | 10445                 | 805 | 11250 |

Pearson chi2(2) = 3.8728 Pr = 0.144

tab RELIGION U5DIED, chi2

| religion    | Under-five Child died |     | Total |
|-------------|-----------------------|-----|-------|
|             | no                    | yes |       |
| Orthodox    | 3249                  | 238 | 3487  |
| Catholic    | 95                    | 11  | 106   |
| Protestant  | 1990                  | 144 | 2134  |
| Muslim      | 4896                  | 391 | 5287  |
| Traditional | 87                    | 9   | 96    |
| Other       | 128                   | 12  | 140   |
| total       | 10445                 | 805 | 11250 |

Pearson chi2(5) = 4.3547 Pr = 0.500

tab NHM U5DIED, chi2

| Number of household members | Under-five Child died |     | Total |
|-----------------------------|-----------------------|-----|-------|
|                             | no                    | yes |       |
| 1-5                         | 4558                  | 457 | 5015  |
| 6-10                        | 5448                  | 324 | 5772  |
| 11-15                       | 377                   | 19  | 396   |
| 16 and above                | 62                    | 5   | 67    |
| total                       | 10445                 | 805 | 11250 |

Pearson chi2(3) = 52.9011 Pr = 0.000

tab SEXH U5DIED, chi2

| Sex of household head | Under-five Child died |     | Total |
|-----------------------|-----------------------|-----|-------|
|                       | No                    | yes |       |
| Male                  | 8516                  | 650 | 9166  |
| Female                | 1929                  | 155 | 2084  |
| total                 | 10445                 | 805 | 11250 |

Pearson chi2(1) = 0.3063 Pr = 0.580

tab WINDEX U5DIED, chi2

| Wealth index | Under-five Child died |     | Total |
|--------------|-----------------------|-----|-------|
|              | no                    | yes |       |
| Poorest      | 3213                  | 287 | 3500  |
| Poorer       | 1895                  | 155 | 2050  |
| Middle       | 1691                  | 127 | 1818  |
| Richer       | 1688                  | 117 | 1805  |
| Richest      | 1958                  | 119 | 2077  |
| total        | 10445                 | 805 | 11250 |

Pearson  $\chi^2(4) = 13.9243$  Pr = 0.008

tab TCHBORN U5DIED,  $\chi^2$

| Total children ever born | Under-five Child died |     | Total |
|--------------------------|-----------------------|-----|-------|
|                          | no                    | yes |       |
| 1-4                      | 6202                  | 431 | 6633  |
| 5-8                      | 3490                  | 289 | 3779  |
| 9 and above              | 753                   | 85  | 838   |
| total                    | 10445                 | 805 | 11250 |

Pearson  $\chi^2(2) = 16.9552$  Pr = 0.000

tab AGEFIRST U5DIED,  $\chi^2$

| Age of mother at first birth | Under-five Child died |     | total |
|------------------------------|-----------------------|-----|-------|
|                              | no                    | yes |       |
| 9-14                         | 795                   | 69  | 864   |
| 15-19                        | 5606                  | 462 | 6068  |
| 20-24                        | 3244                  | 225 | 3469  |
| 25-29                        | 652                   | 36  | 688   |
| 30 and above                 | 148                   | 13  | 161   |
| Total                        | 10445                 | 805 | 11250 |

Pearson  $\chi^2(4) = 9.1893$  Pr = 0.057

tab CMSTATUS U5DIED,  $\chi^2$

| Current marital status    | Under-five Child died |     | Toatal |
|---------------------------|-----------------------|-----|--------|
|                           | No                    | yes |        |
| Married                   | 9174                  | 723 | 9897   |
| Living with partner       | 535                   | 32  | 567    |
| Widowed                   | 192                   | 15  | 207    |
| Divorced                  | 363                   | 29  | 392    |
| No longer living together | 181                   | 6   | 187    |
| total                     | 10445                 | 805 | 11250  |

Pearson  $\chi^2(4) = 6.7068$  Pr = 0.152

tab HEDU U5DIED,  $\chi^2$

| Husband educational level | Under-five Child died |     | Total |
|---------------------------|-----------------------|-----|-------|
|                           | No                    | yes |       |
| No education              | 5362                  | 453 | 5815  |
| Primary                   | 3808                  | 274 | 4082  |
| Secondary                 | 736                   | 44  | 780   |
| Higher                    | 442                   | 27  | 469   |
| Don't know                | 97                    | 7   | 104   |
| total                     | 10445                 | 805 | 11250 |

Pearson  $\chi^2(4) = 8.8344$  Pr = 0.065

tab HOCCUP U5DIED,  $\chi^2$

| Husband occupation    | Under-five Child died |     | Total |
|-----------------------|-----------------------|-----|-------|
|                       | No                    | Yes |       |
| Jobless               | 1880                  | 116 | 1996  |
| Agricultural worker   | 7464                  | 617 | 8081  |
| Trade worker          | 417                   | 23  | 440   |
| Operator or assembler | 202                   | 9   | 211   |
| Lobourer              | 469                   | 40  | 509   |
| Other                 | 13                    | 0   | 13    |
| total                 | 10445                 | 805 | 11250 |

Pearson  $\chi^2(6) = 14.7208$  Pr = 0.023

tab MOCCUP U5DIED,  $\chi^2$

| Mother's occupation    | Under-five Child died |     | Total |
|------------------------|-----------------------|-----|-------|
|                        | No                    | Yes |       |
| Not working            | 5721                  | 430 | 6151  |
| Professional/technical | 167                   | 5   | 172   |
| Sales/services         | 1641                  | 115 | 1756  |
| Agricultural           | 2163                  | 186 | 2349  |
| Manual                 | 753                   | 69  | 822   |
| total                  | 10445                 | 805 | 11250 |

Pearson  $\chi^2(4) = 9.8523$  Pr = 0.043

tab BORD U5DIED,  $\chi^2$

| Birth order number | Under-five Child died |     | Total |
|--------------------|-----------------------|-----|-------|
|                    | No                    | Yes |       |
| 1-3                | 5378                  | 407 | 5785  |
| 4-6                | 3307                  | 255 | 3562  |
| 7-9                | 1431                  | 120 | 1551  |
| 10 and above       | 329                   | 23  | 352   |
| total              | 10445                 | 805 | 11250 |

Pearson  $\chi^2(3) = 1.1195$  Pr = 0.772

tab SEX U5DIED,  $\chi^2$

| Sex of child | Under-five Child died |     | Total |
|--------------|-----------------------|-----|-------|
|              | No                    | Yes |       |
| Male         | 5340                  | 442 | 5782  |
| Female       | 5105                  | 363 | 5468  |
| Total        | 10445                 | 805 | 11250 |

Pearson  $\chi^2(1) = 4.2793$  Pr = 0.039

tab MBFEED U5DIED,  $\chi^2$

| Months of breast feeding | Under-five Child died |     | Total |
|--------------------------|-----------------------|-----|-------|
|                          | No                    | Yes |       |
| 0-6 months               | 1336                  | 1   | 1337  |
| 7-12 months              | 938                   | 0   | 938   |
| 13-18 months             | 868                   | 2   | 870   |
| 19-59 months             | 1486                  | 3   | 1489  |

|                                         |       |     |       |
|-----------------------------------------|-------|-----|-------|
| Ever breastfed, not currently breastfed | 5646  | 542 | 6188  |
| Never breast fed                        | 171   | 257 | 428   |
| Total                                   | 10445 | 805 | 11250 |

Pearson  $\chi^2(5) = 2.2e+03$  Pr = 0.000

tab AGEM U5DIED,  $\chi^2$

| Age of mother | Under-five Child died |     | Total |
|---------------|-----------------------|-----|-------|
|               | No                    | Yes |       |
| 15-19         | 444                   | 50  | 494   |
| 20-24         | 2048                  | 187 | 2235  |
| 25-29         | 2179                  | 140 | 2319  |
| 30-34         | 1574                  | 140 | 1714  |
| 35-39         | 1574                  | 104 | 1714  |
| 40-44         | 664                   | 50  | 714   |
| 45-49         | 223                   | 23  | 246   |
| Total         | 10445                 | 805 | 11250 |

Pearson  $\chi^2(6) = 26.2794$  Pr = 0.000

tab REGION U5DIED,  $\chi^2$

| Region            | Under-five Child died |     | Total |
|-------------------|-----------------------|-----|-------|
|                   | No                    | Yes |       |
| Tigray            | 1078                  | 78  | 1156  |
| Afar              | 1003                  | 95  | 1098  |
| Amhara            | 1163                  | 88  | 1251  |
| Oromia            | 1607                  | 122 | 1729  |
| Somali            | 907                   | 73  | 980   |
| Benishangul-Gumuz | 903                   | 91  | 994   |
| SNNP              | 1449                  | 116 | 1565  |
| Gambela           | 737                   | 57  | 794   |
| Harari            | 598                   | 43  | 641   |
| Addis Ababa       | 361                   | 11  | 372   |
| Diredawa          | 639                   | 31  | 670   |
| Toatl             | 10445                 | 805 | 11250 |

Pearson  $\chi^2(10) = 26.8208$  Pr = 0.003

Appendix B: Summary of descriptive statistics for Outliers and Influential observations diagnostics

Descriptive Statistics

|                                       | N     | Minimum   | Maximum   |
|---------------------------------------|-------|-----------|-----------|
| Predicted probability                 | 1125  | 0.0000174 | 0.8893694 |
| Analog of Cook's influence statistics | 11250 | 0.000074  | 0.895731  |
| Leverage value                        | 11250 | 0.0000273 | 0.1508712 |
| DFBETA for AGEM(15-19)                | 11250 | 0.003968  | 0.1917447 |
| DFBETA for AGEM(20-24)                | 11250 | 0.000339  | 0.7456798 |
| DFBETA for AGEM(25-29)                | 11250 | 0.0003066 | 0.1495911 |
| DFBETA for AGEM(30-34)                | 11250 | 0.0000959 | 0.445777  |
| DFBETA for AGEM(35-39)                | 11250 | 0.0005014 | 0.1909302 |
| DFBETA for AGEM(40-44)                | 11250 | 0.0005679 | 0.6689607 |
| DFBETA for AGEM(45-49)                | 11250 | 0.0004444 | 0.0006065 |
| DFBETA for REGION                     |       |           |           |
| DFBETA for Tigray                     | 11250 | 0.0024582 | 0.1806753 |
| DFBETA for Amhara                     | 11250 | 0.0001988 | 0.1495911 |



|                                   |       |           |            |
|-----------------------------------|-------|-----------|------------|
| DFBETA for Oromia                 | 11250 | 0.0003161 | 0.2688236  |
| DFBETA for Somali                 | 11250 | 0.0001789 | 0.6899701  |
| DFBETA for Benishangul            | 11250 | 0.0006058 | 0.714471   |
| DFBETA for SNNP                   | 11250 | 0.0002063 | 0.8415708  |
| DFBETA for Gambela                | 11250 | 0.0006256 | 0.6592283  |
| DFBETA for Harari                 | 11250 | 0.0002529 | 0.449236   |
| DFBETA for Afar                   | 11250 | 0.000548  | 0.6098817  |
| DFBETA for Diredawa               | 11250 | 0.0000195 | 0.611441   |
| DFBETA for Addis Ababa            | 11250 | 0.0000174 | 0.6973433  |
| DFBETA for NHM                    |       |           |            |
| DFBETA for NHM(1-5)               | 11250 | 0.0003406 | 0.69734333 |
| DFBETA for NHM(6-10)              | 11250 | 0.0001947 | 0.0619285  |
| DFBETA for NHM(11-15)             | 11250 | 0.0001886 | 0.0370294  |
| DFBETA for NHM(16 and above)      | 11250 | 0.000586  | 0.6562303  |
| DFBETA for TCHBORN                |       |           |            |
| DFBETA for TCHBORN(1-4)           | 11250 | 0.0003406 | 0.6973433  |
| DFBETA for TCHBORN(5-8)           | 11250 | 0.0006874 | 0.0981652  |
| DFBETA for TCHBORN(9 above)       | 11250 | 0.001435  | 0.6447405  |
| DFBETA for AGEFIRST               |       |           |            |
| DFBETA for AGEFIRST(15-19)        | 11250 | 0.0004827 | 0.077487   |
| DFBETA for AGEFIRST(20-24)        | 11250 | 0.0002063 | 0.1495911  |
| DFBETA for AGEFIRST(25-29)        | 11250 | 0.0006065 | 0.6973433  |
| DFBETA for AGEFIRST(30 and above) | 11250 | 0.0005028 | 0.1888547  |
| DFBETA for BORD                   |       |           |            |

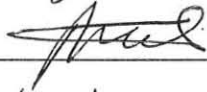
|                                    |       |           |           |
|------------------------------------|-------|-----------|-----------|
| DFBETA for BORD(1-3)               | 11250 | 0.0003406 | 0.0832011 |
| DFBETA for BORD(4-6)               | 11250 | 0.0004827 | 0.0981652 |
| DFBETA for BORD(7-9)               | 11250 | 0.0003813 | 0.0755653 |
| DFBETA for MBFEED                  |       |           |           |
| DFBETA for MBFEED(0-6)             | 11250 | 0.0002378 | 0.0007698 |
| DFBETA for MBFEED(7-12)            | 11250 | 0.0049236 | 0.013902  |
| DFBETA for MBFEED(13-18)           | 11250 | 0.0015572 | 0.0016765 |
| DFBETA for MBFEED(19-59)           | 11250 | 0.0004534 | 0.0017616 |
| DFBETA for MBFEED(ever breastfed)  | 11250 | 0.0538575 | 0.2696815 |
| DFBETA for MBFEED(never breastfed) | 11250 | 0.3615084 | 0.8957312 |
| DFBETA for MEDU                    |       |           |           |
| DFBETA for MEDU(no education)      | 11250 | 0.0002063 | 0.7456798 |
| DFBETA for MEDU(primary)           | 11250 | 0.0004864 | 0.6973433 |
| DFBETA for MEDU(secondary)         | 11250 | 0.0000641 | 0.0239362 |
| DFBETA for MEDU(higher)            | 11250 | 0.0000353 | 0.213977  |

### DECLARATION

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

Declare by

Name Ashagne Sharen

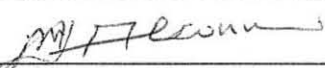
Signature 

Date 26/06/2013

Place of submission: Department of statistics, college of natural science, Addis Ababa University.

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

Name Mekonnen Tadesse

Signature 

Date 26/06/2013