55

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACTORS AFFECTING WEIGHT-FOR-AGE STATUS OF CHILDREN IN ETHIOPIA

GETACHEW TADESSE

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 University

Addis Ababa, Ethiopia

June, 2013

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

FACTORS AFFECTING WEIGHT-FOR-AGE STATUS OF CHILDREN IN ETHIOPIA

GETACHEW TADESSE

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 University

Addis Ababa, Ethiopia

June, 2013

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by Getachew Tadesse, entitled: ***factors affecting weight-for-age status of children in Ethiopia*** and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Statistics complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner Gashaw Wonaheka Signature ______ Date 26/06/2013

Examiner M. K. Sharma Signature M.K.Sh Date 26/06/13

Advisor Mekonnen Tadesse Signature ______ Date 26/06/2013

____

Chair of Department or Graduate Program Coordinator

ABSTRACT

Factors affecting weight-for-age status children in Ethiopia

Getachew Tadesse

Addis Ababa University, 2013

Weight-for-age is a nutritional status indicator of malnutrition (acute or chronic) based on the principle that a child has an expected weight for his/her age and measures the general nutritional status of children. According to the 2011 EDHS report by the Central CSA of Ethiopia, nationally, 29 percent of children under the age of five are underweight, and 9 percent of children are severely underweight. The main objective of this study was to identify and describe factors affecting the weight-for-age status of children in Ethiopia. The study was based on the 2011 EDHS conducted by CSA. Descriptive, standard logistic regression and multilevel logistic regression analyses were used. The logistic regression analysis was carried out to investigate the effect of predicator variables on the weight-for-age status of a child. Accordingly, Age of child, Source of drinking water, Types of toilet facilities, Mother's educational level, Employment status of mothers, Wealth index, Place of residence, experiencing diarrhea, Fever and Cough in the two weeks before the survey, Educational level of partner and Geographical region were found to be important determinants of weight-for-age status of under-five children in Ethiopia. Moreover, the variables that were reported to be significant in the ordinary logistic regression analysis were also found to be significant in the multilevel logistic regression analysis. The effect of most of these significant variables was the same for each region in Ethiopia but the effect of Employment status of mothers and economic status (wealth index) was not the same for all regions in Ethiopia.

ACKNOWLEDGEMENT

First, and foremost, I thank God for giving me the opportunity to pursue my graduate study at the Department of Statistics, Addis Ababa University.

I owe the deepest gratitude to Mekonnen Tadesse (Assistant Prof.), my thesis advisor for his invaluable and constructive comments and suggestions throughout my study. Without his professional assistance and guidance this study would not be realized.

My special thanks also go to Prof. Eshetu Wencheko and Dr. Melaku Wakuma who helped me to register and complete my thesis, Department of Statistics AAU, and to my colleagues for all the encouragements and support that they provided me during my study years.

Finally and most importantly, my heart-felt thanks go to all members of my family especially to my brother Tarekegn Tadesse and my sister Ayelech Tadesse for all the encouragement and support throughout my study.

TABLE OF CONTENT

LIST OF ABBREVIATIONS

AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
CSA	Central Statistical Agency
DHS	Demographic and Health Survey
EDHS	Ethiopian Demographic and Health Survey
FAO	Food and Agricultural Organization
LR	likelihood ratio
MDG	Millennium Development Goals
ML	Maximum Likelihood
NCHS	National Center for Health Statistics
S.E	Standard Error
SC	Schwarz Criteria
SD	Standard Deviation
UN	United Nations
UNICEF	United Nations Children's Fund
Var	Variance
WHO	World Health Organization

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1 Research background

Adequate nutrition is essential for proper growth and physical development from conception to adulthood to ensure optimal working capacity, normal reproductive performance and adequacy of immune mechanism which provides resistance to infections (Bender, 1997). Under-nourishment due to protein energy malnutrition significantly increases susceptibility to major infectious diseases in low-income countries, particularly in children (Schaible and Kaufmann 2007, Ambrus and Ambrus 2004, Woodward B. 1998). Malnutrition is responsible, directly or indirectly, for 54% of the 10.8 million deaths per year in children under-five and contributes to every second death (53%) associated with infectious diseases among children under-five years of age in developing countries (Schaible and Kaufmann (2007)). Throughout the developing world, 1 out of every 4 children roughly 146 million children under the age of five is underweight (State of the World's children 2007).

The highest levels of underweight prevalence are found in South Asia, where almost half (46%) of all children under-five are underweight, followed by Sub- Saharan Africa, with more than one-quarter (28%) of all children under-five being underweight. Three countries in South Asia India, Bangladesh and Pakistan–home to merely 29% of the emerging worlds under-five populations, accommodate 50% of the world's underweight children (UNICEF, 1998).

Malnutrition stems from a number of causes but at most immediate level it occurs as a consequence of inadequate dietary intake and disease, which usually occur in combination. However, a number of cultural and economic elements are also involved (UNICEF, 2006).

Child growth is internationally recognized as an important public health indicator for monitoring nutritional status and health in populations. A number of studies have established the association between increasing severity of anthropometric deficits and mortality (De Onis and Blössner, 2003). Of the various anthropometric indices that can be used to assess child growth status, three are more widely used: height-for-age portrays performance in terms of linear growth, and essentially measures long-term growth faltering (stunting); weight-for-height reflects body proportion, or the harmony of growth, and is particularly sensitive to acute growth disturbances (wasting); and weight-for-age represents a convenient synthesis of both linear growth and body proportion (underweight) (WHO Working Group, 1986).

Nutritional problems exist all over the world and are particularly noticeable among African children at pre-school level. Among African children there is chronic malnutrition, which is not serious enough to cause them to suffer from but results in them being both physically and mentally stunted, wasted and underweight. Children, especially those under 5 years, are the most vulnerable group in an environment that does not guarantee adequate food and protection. There is a high nutritional requirement for children under 5 years of age because they are undergoing a period of rapid growth. Good nutrition at this age is vital, as unavailability of certain key nutrients for a given age could result in physical and mental retardation that may be irreversible. In most cases, African children are more affected because of their unfavorable social and economical conditions. (Rossouw, 1996:29)

Ethiopia is one of the least developed countries in the world as measured by the (Purchasing Power Parity) GNP per capita or Human Development Index (UNDP, 2002; World Bank, 2004), with about 44 percent of the population living below the poverty line (MoFED, 2002a). In 2001, the under-five mortality rate in Ethiopia was about 17 percent (World Bank, 2004). As is the case in most developing countries, most of the causes of under-five mortality in Ethiopia are related to acute lower respiratory tract infections (ARI), malaria, diarrhea, malnutrition and measles (MoH, 2002).

Malnutrition is alone responsible for over 50 percent of deaths of under-fives in developing countries, making it one of the most important public health problems (Pelletier and Frogillo 1994; Murray and Lopez 1997). The consequences of child malnutrition include poor physical development and limited intellectual abilities which can have negative impacts on economic growth because they diminish the working capacity of these children when they reach adulthood (Girma and Genebo, 2002; Alderman *et al.*, 2004).

According to the Central Statistics Authority's Demographic and Health Survey (CSA, 2000), more than 50 percent of children under-five were stunted, some 11 percent were wasted, and 47 percent were underweight (low weight-for-age). This compares poorly with the sub-Sahara average incidence of underweight and stunted children in the 1990s, which was 33 percent and 39 percent respectively (Christiaensen and Alderman, 2004).

Ethiopia is one of the developing countries where the problem of malnutrition is still prevailing. Being the second most populated country in Africa, the country has to still struggle to combat poverty and work its way toward food security through, largely its rain dependent farming and livestock production. Most national surveys and studies conducted in recent years affirm the presence of malnutrition to the degree that it is a major Public Health Problem. Though child mortality is declining, around 270,000 under-five deaths are still happening due to malnutrition. The 2005 Ethiopian Demographic and Health Survey (EDHS) reported remarkably high rates of underweight (38%) prevalence in the country.

Anthropometric measurements (weight, height and age) can be used to assess the degree of malnutrition among population groups and help, for instance, to set priorities of the food targeting policies to the severely malnourished groups. Nutritional indices in the case of children are sensitive indicators and signal serious problems that might require an immediate policy response. This malnutrition of children has negative effect on the development of the country. In this study, we will only be focusing on underweight as an indicator of malnutrition in our study population.

1.2 Statement of the problem

Almost one in every ten babies born in Ethiopia (97 per 1000) does not survive to celebrate its first birthday, and one in every six children dies before its fifth birthday. As a result, it will be challenging to reach the child survival Millennium Development Goals (reducing child mortality by 3/4) with the current pace of mortality reduction (WB and MoH, 2005).

High malnutrition rates in Ethiopia pose a significant obstacle to achieving better child health outcomes. Malnutrition among under-five children is a chronic problem in developing countries like Ethiopia. Worldwide, ten and a half million children of age under-five die every year, with 98% of these deaths reported to occur in developing countries (UNICEF, 2007). In recognition of the burden of malnutrition among under-five children, four of the eight United Nations' Millennium Development Goals (MDGs) are specifically directed towards improving child health outcomes in developing countries. In particular, a reduction in the mortality of children is a key MDG, and a reduction in malnourishment among children is an important indicator of progress towards that goal. Even though the problem of child malnutrition in Ethiopia has been sufficiently documented, the reasons behind it are still poorly understood. There are also inconsistencies across studies regarding the determinant factors behind childhood weight-for-age status of children. Therefore, this study attempted to investigate the major socio-economic, demographic, health and environmental determinants of weight-for-age status of children in Ethiopia.

1.3 Objective of the study

The general objective of this study is to identify factors affecting weight-for-age status of under-five children in Ethiopia.

The specific objectives of this study are:

1. Examine the extent of the variation in weight-for-age status of children within and between regions of Ethiopia.

2. To identify the factors that may explain the variation in childhood weight-for-age status between regions of Ethiopia.

3. To estimate the level of predictors for weight-for-age among under five children in Ethiopia.

1.4 Significance of the study

This research is not an attempt to find the medication or cure for factors that affect the children health status Instead, it is an attempt to examine the prevalence pattern and factors that affect weight-for-age status of children in Ethiopia. Through this research an analysis of children nutritional status in relation to demographic, socio-economic, and feeding practice will be presented.

The findings could be helpful for policy making, monitoring and evaluation activities of the government and different concerned agencies. The result could be useful to guide the end users (governmental and non-governmental organizations) to develop nutrition programs and set appropriate plans to tackle the existing health and nutrition problems.

1.5 Limitation of the study

The data used here being secondary may have a number of limitations on the outcome of this study.

- There are many factors associated with underweight as indicated by different studies in different countries, this study is undertaken to explore a few of the socio-economic, demographic, health and environmental factors in Ethiopia.
- Absence of periodical reports, and approved resources.

1.6 organization of the paper

The second chapter is a review of literature related to factors affecting weight-for-age status of under-five children in Ethiopia and the rest of the world. The third chapter describes the source of data, variables of the study and the methodology used for analysis. The fourth chapter presents the statistical data analysis while the last chapter provides discussion, conclusion and recommendations.

CHAPTER TWO: LITERATURE REVIEW

This section briefly reviews studies related to factors affecting children's nutritional status. Variables related to household economic welfare, household composition and education, access to services are considered in turn.

2.1. Anthropometric Indicators

Changes in body dimensions reflect the overall health and welfare of individuals and populations. Anthropometry is used to assess and predict performance, health and survival of individuals and reflect the economic and social well being of populations. Anthropometry is a widely used, inexpensive and non-invasive measure of the general nutritional status of an individual or a population group. Recent studies have demonstrated the applications of anthropometry to include the prediction of who will benefit from interventions, identifying social and economic inequity and evaluating responses to interventions. Anthropometry can be used for various purposes, depending on the anthropometric indicators selected.

The four building blocks or measures used to undertake anthropometric assessment are: age, sex length (or height) and weight.

Each of these variables provides one piece of information about a person. When they are used together they can provide important information about a person's nutritional status. The actual measurement of age, weight and height of children requires specific equipment and techniques which are described later. When two of these variables are used together they are called an index. Among the three indices commonly used in assessing the nutritional status of children, one is Weight-for-age.

Weight-for-age: Low weight-for-age index identifies the condition of being underweight, for a specific age. The advantage of this index is that it reflects both past (chronic) and present (acute) under nutrition (although it is unable to distinguish between the two).

Underweight: Underweight, based on weight-for-age, is a composite measure of stunting and wasting and is recommended as the indicator to assess changes in the magnitude of malnutrition over time.

2.1.1 Malnutrition

The effects of malnutrition on human performance, health and survival have been the subject of extensive research for several decades and studies show that malnutrition affects physical growth, morbidity, mortality, cognitive development, reproduction, and physical work capacity (Pelletier and Frongillo, 2002). Malnutrition is an underlying factor in many diseases in both children and adults, and it contributes greatly to the disability-adjusted life year's worldwide (Murray and Lopez, 1996). Malnutrition is particularly prevalent in developing countries, where it affects one out of every three preschool-age children (United Nations Sub-Committee on Nutrition, 2004).

A well-nourished child is one with access to adequate food supply, care and health. Such a child will have weight and height measurements that compare very well with the standard normal distribution of heights (H) and weights (W) of healthy children of the same age and sex. Thus, the best way to evaluate the nutritional status and overall health of a child is to compare the child's growth indices with the set cut-off points in the standard normal distribution of well nourished children that are associated with adequate growth (De Onis M. et al., 2003).

Factors that contribute to malnutrition are many and varied. The primary determinants of malnutrition, as conceptualized by several authors relate to unsatisfactory food intake, severe and repeated infections, or a combination of the two (UNICEF, 1998, Rowland. et al., 1988, Schroeder and Brown, 1994). The interactions of these conditions with the nutritional status and overall health of the child - and by extension - of the populations in which the child is raised have been shown in the UNICEF Conceptual framework of child survival (UNICEF, 1998). Briefly, the model characterizes the correlates of malnutrition as factors that impair access to food, maternal and child care, and health care. It is these very factors that impact the growth of children. Consequently, the assessment of children's growth is a suitable indicator for investigating the wellbeing of children, and as well as for examining households' access to food, health and care (De Onis M. et al., 2003, UNICEF, 1998).

2.1.2 Nutritional assessment of Underweight

Growth is a process during which the fertilized zygote turns into mature human. Growth is a complicated process affected by various factors including nutrition, diseases, genetic and racial factors, etc (Rapaport and Bowlby, 2004). Malnutrition and growth failure are among the most common problems in children and it is also the cause of half of children's deaths in developing countries (Phengxay et al., 2007). Malnutrition is manifested by different patterns such as underweight (Baradaran and Hasani, 2006). The presence of such conditions increases the risk of developing infectious diseases, parasitic infections, immunologic disorders, reduced functionality.

Determination of underweight was performed according to the classification of WHO. The indices of under-nutrition such as underweight were calculated by Z-score using the following formula of the WHO standard. The statistical equation used was (z-score $WA=x_i-\mu$) for calculation of underweight. The cases were categorized as mild, intermediate and severe based on WHO classification protocol.

In this manner, children with Z-score at (-1SD to -2SD) from the mean were considered as mildly underweight, at (-2SD to -3SD) as intermediately underweight, and at <-3SD as severely underweight.

2.2 Household economic welfare

According to Christiaensen and Alderman (2004), sustained income growth of 2.5 percent per adult over a fifteen-year period could be associated with a 3-6 percent decline in chronic child malnutrition in Ethiopia. They argued, however, that income growth alone might not be sufficient to alleviate child malnutrition. Similarly, Glewwe *et al.* (2002) observed that, due to rapid economic growth in Vietnam since 1986, a dramatic decline in poverty and child stunting, wasting and underweight was achieved. Based on their empirical investigation, they concluded that growth in household income, although not very large, had a positive impact on child nutrition in Vietnam during the 1990s. They also noted that, over time, child stunting, wasting and underweight declined within each quintile even after adjustment for change in income was made, which suggests that there are other factors, in addition to income growth, which led to improvements in child nutrition.

The better off households have better access to food and higher cash incomes than poor households, allowing them a quality diet, better access to medical care and more money to spend on essential non-food items such as schooling, clothing and hygiene products (Stalin p. et al., 2013). The study showed that the prevalence of underweight among under-five children who belongs to class poorest and class poor socio-economic status was 47.2% and 63.4% respectively. Under-five children belonging to class medium socio-economic status were underweight (46.7%).

2.3 Household factors

A growing number of studies recognize the important role that household composition and parental levels of education play in shaping children's nutritional outcomes.

2.3.1 Household composition

The evidence on household composition and child malnutrition is mixed. Christiaensen and Alderman (2004) found that larger family size results in underweight among children in Ethiopia. They argue that economies of scale in time for childcare and expenditure can be enjoyed in large families and that children benefit from parents' accumulated experiences in care of young children. However, Desai's (1995) cross-country study found that the nutrition of a child aged between 6 and 36 months is likely to be negatively affected by the presence of a sibling aged 0-5 years as siblings of a similar age are likely to compete for parental resources. Having siblings of 6-12 years has a less clear effect, which is sometimes negative. The presence of siblings of 13-15 years has a positive effect on a child's nutritional status.

Desai argues that the ability for parents to provide adequate nutrition for young children in large families may be enhanced by the existence of extended kinship networks or, alternatively, state subsidies for child-related expenses.

2.3.2 Sex and age of a child

A number of studies in Africa suggest that rates of malnutrition among boys are consistently higher than among girls (Svedberg, 1990; MoFED, 2002a; Glewwe *et al.*, 2002; Sahn and Stifel, 2002; Christiaensen and Alderman, 2004). Sahn and Stifel (2002) present three possible explanations for such a result. The first is that there is a problem with the gender-specific standard for the African population, similar to the inconsistencies noted regarding the weight-for-height curves for young children (<2 years) in a number of African countries (Macfarlane, 1996). The second is that girls are genetically more robust than boys are. There is some evidence that among pre-adolescent children there is an increment of fat in preparation for the growth spurt; this was also true for our sample. This increase of fat is greater for girls than boys (Brown, 2002). The third is that there is greater investment in young girls than boys, although this would be counter-intuitive since girls are widely regarded as a 'poorer investment' since they eventually marry and leave to join their husbands' family. We may also add a fourth possible explanation: that it reflects boys' and girls' gender-specific roles - i.e. girls' better access to food through their roles in cooking, together with a combination of boys' higher energy expenditure and lack of food during the day when they are involved in tasks such as herding animals. Note also that the energy requirement is different for boys and girls. When we see the age, children with previous birth interval 0-23 months and 24-47 months had respectively 1.4 times and 1.2 times higher risk of being under-weight as compared to children with previous birth interval 48 and above months. Babies were very small in size and smaller than average had respectively 3.93 times and 2.23 times higher risk of being underweight than those children who were average or large in size at birth. (Md.Israt Rayhan and Sekander, 2006)

2.3.3 Household education

Various studies have concluded that parental education, especially mothers' education, is a key element in improving children's nutritional status (Moen, 1993; Christiaensen and Alderman, 2004). Glewwe (1999) investigated the mechanisms through which education (schooling) results in better child nutrition or health. Specifically, Glewwe identified three possible pathways:

1) Direct lessons in nutrition and health in schools to future mothers;

2) Gaining nutritional knowledge because of literacy and numeracy skills acquired in school; and 3) exposure to, and familiarity with, modern society through school which facilitates the use of modern medicines. The implication is that schooling is associated with child nutrition only if it can improve mothers' nutritional knowledge. Glewwe (1999) further suggested that such knowledge could be attained outside the classroom. As will be discussed in the concluding section on policy implications, this finding may be of significance for a developing country like Ethiopia, it is possible to improve nutritional knowledge through specific nutrition education programmers' (Christiaensen and Alderman, 2004) without formal schooling. However, the impact of mothers' schooling on child nutrition is not only through nutritional knowledge. An educated mother is likely to have a higher income (which can directly affect her children's health and nutrition) and higher status and power in the household and community (which will put her in a better position to make decisions about her children's needs) (Moen, 1993). The literature also notes that, particularly where the general level of education of the community is low, the

level of education of female and male members of the household could be particularly important in indirectly influencing child nutritional status (Basu and Foster, 1998; Gibson, 2001). More specifically, Christiaensen and Alderman (2004) found that the effect of maternal education is about twice as important as that of paternal education. The education of fathers had an effect on children to being underweight. The study conducted in Pakistan, the logistic regression analysis showed that father's education and prevalence of underweight were inversely related. Risk of underweight were 0.98 and 0.70 times lower for children of fathers attended primary and secondary level respectively, comparing to the children of illiterate fathers, Md.israt Rayhan and M.sekander (2006).

2.4 Access to services

2.4.1 Water and sanitation

A household's access to facilities is likely to be correlated with community characteristics. Households living in wealthier communities might have a relatively healthy environment, which implies better sanitation facilities, access to clean water and healthcare facilities (Glewwe *et al.*, 2002). Water and sanitation play a particularly important role in child nutrition due to their impact on diarrheal diseases.

Christiaensen and Alderman (2004) noted that in Ethiopia, 14 percent of households in urban communities get water from their own tap and 3 percent have flush toilets. However, no households in rural areas had piped water in their home and only 1 percent had flush toilet facilities. Their findings indicate a significant and positive impact of these facilities on child nutrition. Glewwe *et al.* (2002) considered the impact of clean water, a sanitary toilet and electricity on child health in Vietnam. They found that, when the three variables are included separately, the lack of a sanitary toilet and electricity had significant negative effects on child health while clean water had no significant effect. When the three variables were added simultaneously, electricity lost its statistical significance while a sanitary toilet remained significant. They concluded that policies directed at improving sanitary toilets in a community would have better health outcomes than either water or electricity alone.

Esrey (1996) and Esrey *et al.* (1992) found in a multi-country study that improved sanitation had a substantial effect on nutritional status, while improvement in water supply enhances these effects. They also note that when considered separately, water supply has a smaller effect on nutritional status than sanitation.

2.4.2 Access to health services

Access to health services is the main determinant of whether the public is going to utilize the services.

However, one also has to consider the quality and prices associated with the service, although they obviously have little relevance if access is limited (MoFED, 2002a; 2002b). Access to health services is expected to influence nutritional status, as children without access to such services are more likely to be malnourished, reflected through weight loss which is associated with untreated diarrhea and other infectious diseases.

One of the important measures of access to health facilities used in the literature is the distance between the household's home and the nearest healthcare facility (MoFED, 2002a; Roberts, 2003; Christiaensen and Alderman, 2004). This indicator, a proxy for availability of healthcare, which affects healthcare choices, is widely used in malnutrition studies (Christiaensen and Alderman, 2004), and is particularly useful for a country like

Ethiopia where transportation networks are poor in terms of coverage, road quality and safety (MoFED, 2002a). The average distance from a household to the nearest health centre, as reported by MoFED (2002a) and based on the welfare monitoring survey of 1999-2000. However, Christiaensen and Alderman (2004) found that after controlling for a number of other determinants of child malnutrition, distance to the nearest health centre is not a significant determinant. They suggest that this lack of explanatory power of the distance variable might be explained by the fact that the proxy for access to health centre does not capture the quality of healthcare provided, which is as important as distance.

CHAPTER THREE: DATA AND METHODOLOGY

3.1 Data source

The sources of data were the 2011 Ethiopia Demographic and Health Survey (EDHS) which is obtained from the Central Statistical Agency (CSA). It is the third major survey designed to provide estimates for the health and demographic variables of interest for the following domains: Ethiopia as a whole, urban and rural areas of Ethiopia (each as a separate domain), and all geographic areas (nine regions namely: Tigray, Affar, Amhara, Oromiya, Somali, Benishangul-Gumuz, Southern Nations, Nationalities and Peoples (SNNP), Gambela and Harari regional states and two city administrations: Addis Ababa and Dire Dawa). In the 2011 EDHS a representative sample of approximately 17,018 households from 624 clusters was selected. The sample was selected in two stages. In the first stage, 624 clusters (187 urban and 437 rural) were selected from the list of Enumeration Areas (EA).

From 11,654 under-five children, only 9,526 are measured for anthropometric measurements weight-for-age. Thus, the analysis presented in this study on weight-for-age status is based on the under-five children with complete anthropometric measurements.

3.2 Variables of the Study

The dependent and independent variables that were considered to affect the children health status were selected based on findings of available similar studies and the available data on the subject. As demonstrated in the literature review, socio-economic, demographic, health and environmental characteristics are considered as important determinants of weight-for-age status in under-five children.

3.2.1 The Response variable

The response variable weight-for-age status measurement was expressed in Standard Deviation (SD) units (Z-score) from the median of the reference population. Children with a measurement of <-2 SD units from the median of the reference population were considered to have less weight for their age (underweight) and children with measurement of (≥-2SD) units from the median of the reference population were classified as not underweight. In this study, the measurement weight-for-age was calculated using the new Child Growth Standards released by the World Health Organization on April 27, 2006. The new Standards are the result of an intensive study initiated by WHO in 1997 to develop a new international standard for assessing the physical growth nutritional status and motor development in all children from birth to age five.

To determine the weight-for-age status of children, the dependent variable was expressed as a dichotomous variable. In view of this, the response variable, weight-for-age status the was measured as a dichotomous variables as follows.

$$Y = \begin{cases} 0, not\ underweight\ \geq -2SD\ \ (-2\ \ Z - score) from\ the\ median\ of\ reference \\ 1, underweight < -2SD\ \ (-2\ \ Z - score) from\ the\ median\ of\ reference \end{cases}$$

Where: Y represent the weight-for-age status.

3.2.2 Explanatory variables

Based on the reviewed literatures, some of the common predictors that are expected to influence the weight-for-age status of under-five children in Ethiopia relevant for our analyses were recorded as given below. The predictor variables to be studied as determinants of weight-for-age status were grouped into three major factors: socio-economic, demographic and health and environmental factors.

Socio-Economic Characteristics

Among the socioeconomic characteristics that may influence the status of weight-for-age of under-five children, the following factors were included: mother's level of education, employment status of mothers, employment status of partners, level of education of partners, household income, place of residence and geographical region.

Demographic Characteristics

Demographic characteristics included were age, sex, birth order and birth interval of the child.

Health and Environmental Characteristics

Health and environmental characteristics that may influence weight-for-age status of under-five children considered in this study were diarrhea, fever, cough, water supply source and toilet facilities.

Table 3.1 Description of the Variables and coding

- **Response Variable**

Variables	Representation of variable	Factor categories
Underweight	Y	0, not underweight 1, underweight

- **Explanatory Variables(factors)**

1. **Demographic variables**

Variables	Categories
Sex of child(SC)	0 = Female 1 = Male
Age of child (AC)	0= 0-5 months 1= 6-11 months 2= 12-23 months 3= 24-35 months 4= 36-47 months 5= 48-59 months
Birth order of the Child (BOC)	0= 1 1= 2-3 2= 4-5 3= 6+

2. **Socio-Economic Variables**

Variables	Categories
Mother's Education(EDM)	0= No education 1= Primary education, 2= Secondary and above
Employment status of mother (ESM)	0=Not employed 1= Employed
Wealth index(WIN)	0= Poor 1= Medium

	2= Rich
Education of partner (EDP)	0= No education 1= Primary education 2= Secondary and above
Number of household members (NHM)	0=1-4 1=5-9 2= 10 and above
Place of Residence (PR)	0= Rural 1= Urban
Region (REG)	1= Tigray 2= Afar 3= Amhara 4= Oromia 5= Somali 6= Ben-gumuz 7= SNNP 8= Gambela 9= Harari 10 = Dire Dawa 11= Addis Ababa

3. Health and Environmental Characteristics

Variables	Categories
Source of drinking water (SDW)	0= Non improved source (surface water, tanker truck, unprotected well/spring, other's) 1= improved source(piped into dwelling, piped to yard/plot, public tap, protected well/spring, rain water, bottled water)
Had - Diarrhea in the two Weeksbefore the survey? (HDW)	0=No 1= Yes

Hadfever in the two weeks before th- survey?(HFW)	0= No 1= Yes
Had cough in the two weeks before the survey? (HCW)	0= No 1= Yes
Type of toilet facility (TTF)	0= No facilities (bush/field toilet, hanging toilet, others) 1= Have facilities (Pit toilet, Flush/pour toilet)

3.3 The Methodology

3.3.1 The Logistic Regression

Logistic regression is a popular modeling approach when the dependent variable is dichotomous or polytomous. This model allows one to predict outcomes, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Hosmer and Lemeshow (2000) have described logistic regression focusing on its theoretical and applied aspect. Logistic model, as compared to its competitor, the probit model, is less sensitive to outliers and easy to correct a bias (Copas, 1988).

Often the outcome variable in social data is, in general not continuous, instead is binary. In such a case, binary logistic regression is a useful way of describing the relationship between one or more independent variables and a binary outcome variable that has only two possible values. Indeed, a generalized linear model is used for binary logistic regression. The most attractive feature of a logistic regression model is that it neither assumes linearity in the relationship between the covariates and the outcome variable, nor does it require normally distributed variables. It also does not assume homoscedasticity and in general has less stringent requirements than linear regression models. Thus logistic regression is used in a wide range of applications leading to binary dependent data analysis (Agresti, 2002).

Regression methods have an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. When the outcome variable is binary, the logistic regression model can be used mainly for two reasons. The first is from a mathematical point of view; it is extremely flexible and easily used function. The second is, it leads itself to meaningful interpretation.

The study also uses multilevel logistic regression analysis because of the hierarchical nature of the data set. Three models were constructed for the multilevel analysis. The first model, an empty model, was without any explanatory variable i.e. simple component of variance analysis. The second model controlled for the individual-level variables, the third model controlled for community-level variables.

3.3.2 The Multiple Logistic Regression

Multiple logistic regression analysis is used to study the effects of the independent variables on the response variable (weight-for-age status of under-five children in Ethiopia). An important aspect in multiple levels of analysis is the partitioning of unexplained variability over the various levels. The logistic regression can have an arbitrary number of parameters and terms in the model representing qualitative variables, quantitative variables and interaction terms in order to model a dichotomous or categorical outcome variable. When explanatory variables are included to model probabilities, a problem is that probabilities are restricted to the domain 0 and 1whereas a linear effect for an explanatory variable could take the fitted value outside this interval.

Let Y_{nx1} be a dichotomous outcome random vector with two categories (coded 1 and 0). Consider a collection of k independent variables which will be denoted by the variable $\boldsymbol{x}= (x_1, x_{2,...,} x_k)'$. Let the conditional probability that the outcome is present be denoted by $P\ (Y=1|\ \boldsymbol{x}) = P(\boldsymbol{x})$ of weight-for-age.

In the terminology of logistic regression analysis the odds of a success is defined to be the ratio of the probability of a success to the probability of a failure. That is the odds of a success is $\frac{p}{1-p}$ where p is the true success probability.

Then the logit of the multiple logistic regressions is given by the equation

$$logit\left[\frac{p}{1-p}\right] = log\left[e^{g(x)}\right] = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad (3.1)$$

where, g (x) $=\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$

In which case,

$$p(x) = \frac{e^{g(x)}}{1+e^{g(x)}} = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_k x_k}} \qquad (3.2)$$

Where β_i's are the model parameters. In order to estimate these parameters, maximum likelihood-based iteration algorithms can be employed.

Variable Selection: the number of variables that would be included in the model should be of the minimum possible that is parsimonious and deliver optimum information. In this study the variable selection process begins with a univariate analysis of each variable. The test can be used to determine whether systematic relations or associations between each predictor variable with the response variable will be made before the selection of the final model. A univariate logistic regression and a likelihood ratio (LR) chi-square test (for a $2 \times L$ contingency table) will also be employed to examine the importance of each predictor variable to the outcome variable.

Upon completion of the univariate analysis, predictor variables for the multivariate analysis will be selected with a condition that any variable whose univariate test has a p-value less than 0.25 is considered as a candidate for the multivariate model along with all variables of known socio-demographic or economic importance (Hosmer and Lemeshow, 2000).

For multivariate model, a stepwise method in which variables are selected either for inclusion or exclusion from the model in a sequential fashion based solely on statistical criteria will be used. There are two main versions of the stepwise procedure: (a) forward selection with a test for backward elimination and (b) backward elimination followed by a test for forward selection. The study employed stepwise forward likelihood ratio procedure in order to select the list of predictor variables that will have joint impact in influencing the outcome variable.

The final decision on the inclusion of each predictor variable will be made on the examination of the Wald statistic for the variable, and comparing each estimated coefficient of the particular variable on the multivariate regression model with the univariate estimate of the model containing only that predictor. Variables that do not contribute to the model based on these criteria will be eliminated and a new model will be fitted. The LR test will be used to compare the new model with the old model. Moreover, the estimated coefficients for the remaining variables will be compared to those from the full model. In view of this (deletion, refitting and/or verifying) will be performed. Having obtained a model that contains the essential variables, the need to include interaction term in the model will be assessed by creating the appropriate product of the variables in question. Assessment of the significance of each interaction term will be made using a LR test. Interactions that did not contribute to the improvement of the model will be discarded and the model with main-effects will be maintained. (Agresti, 2002)

3.3.2.1 Parameter Estimation

The maximum likelihood and non-iterative weighted least squares are the two most competing estimation methods used in fitting logistic regression model (Hosmer and Lemeshow, 2000, Collet, 1991). When the assumption of normality of the predictors does not hold, the non-iterative weighted least squares method is less efficient (Maddala, 1997). In contrast, the maximum likelihood estimation method is appropriate for estimating the logistic (logit) model parameters due to the less restrictive nature of the underlying assumptions (Hosmer and Lemeshow, 2000). Hence, in this study the maximum likelihood estimation technique will be applied to estimate unknown parameters of the model.

Consider the logistic model, $p_i = \frac{e^{x_i\beta}}{1+e^{x_i\beta}}$ (3.3)

Where,x_i's (i=1, 2... k) are predictor variables

$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2 \ldots \beta_k)'$ is a (k+1) x1 vector of parameters.

Since the observed values of y say, $y_i's$ (i= 1, 2... n) are independently distributed as Bernoulli with parameter p_i,y_i~ Bernoulli (1, p_i),the likelihood function of y is given by:

$$\ell(\boldsymbol{Y},\boldsymbol{\beta}) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} = \prod_{i=1}^{n}\left[\frac{e^{x_i\beta}}{1+e^{x_i\beta}}\right]^{y_i}\left[\frac{1}{1+e^{x_i\beta}}\right]^{1-y_i} \tag{3.4}$$

The objective is then to get an estimator $\widehat{\boldsymbol{\beta}}$= ($\hat{\beta}_0$, $\hat{\beta}_1$..., $\hat{\beta}_k$) of $\boldsymbol{\beta}$ which maximizes the likelihood function expressed in Equation (3.4). Since the likelihood equations are nonlinear in the parameters, the Newton-Raphson iterative maximum likelihood estimation method that expresses $\hat{\beta}$ at the $(u+1)^{th}$ cycle of the iteration is expressed as:

$$\hat{\beta}_{u+1} = \hat{\beta}_u + (X'\hat{V}_uX)^{-1}XR_u \tag{3.5}$$

Where, u= 0, 1, 2, 3 ..., and $\hat{v}$ is a diagonal matrix with its diagonal elements,

$\hat{v}$= $diag[\ \hat{p}\ (1-\hat{p}\ )]$ =Cov(Y).

Finally, $\hat{\beta}$ is the resultant maximum likelihood estimator of β with residual $R = \boldsymbol{Y} - \hat{p}$ (Collet, 1991).

3.3.2.2 Test of Goodness of Fit

Once a model has been developed, we would like to know how effective the model is in describing the outcome variable. This is referred to as goodness-of-fit. In testing the hypothesis that the model fits the data, the two common approaches are Pearson's X^2 and the likelihood-ratio tests; see details in Agresti (2002).

The Hosmer-Lemeshow test is another alternative to check model fit. In this approach, data are divided into 10 groups. From each group, the observed and expected number of events will be computed. Then, the Hosmer-Lemeshow test statistic is given by

$$\hat{C} = \sum_{i=1}^{g} \left(\frac{(O_i - E_i)^2}{V_i}\right) \qquad (3.6)$$

Where $E_i = n P_i$, $V_i = n P_i (1 - P_i)$, g is the number of group, O_i is observed number of events in the i^{th} group, E_i is expected number of events in the i^{th} group, and V_i is a variance correction factor for the i^{th} group. If the observed number of events differs significantly 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.

If the calculated value of the Hosmer and Lemeshow goodness-of-fit test statistic is not significant, the model estimates are adequate to fit the data at an acceptable level.

The likelihood ratio (LR) test, which is defined as $-2[Lo - L_1]$ (where L_0 and L_1 are the maximized log-likelihoods under the null and alternative hypothesis, respectively) will be used to test the null hypothesis that the k-coefficients for the covariates in the model are not important in explaining the response variable against the alternative that at least one of the covariates is important. Under the null hypothesis, the LR is distributed as chi-square and if it is significant, we reject the null hypothesis and conclude that at least one of the k-covariates included in the model are important in explaining the variation in the outcome variable.

Test for a Single Predictor: the separate effects of each predictor variable in explaining the outcome variable will be made by postulating the null hypothesis that: $H_{o:}$ $\beta_i = 0$ against the alternative: H_1: $\beta_i \neq 0$ i = 1, 2… n. The appropriate test statistic is the Wald chi-square statistic given by:

$$W = \frac{\hat{\beta}_i^2}{Var(\hat{\beta}_i)} \tag{3.7}$$

For large sample size this statistic has an approximate chi-square distribution with one degree of freedom. The Wald statistic, however, has some undesirable properties for large coefficients as standard error is inflated lowering the Wald statistic (chi-square) value and leading to type II errors. On the other hand, Agresti (2002) suggested that the likelihood ratio (LR) test is more reliable for small sample sizes than the Wald test. Accordingly, this study will use the Wald and/or LR test statistic to assess the significance of each predictor variable.

3.3.3 Multilevel Analysis

The 2011 EDHS data set used for this study is based on multistage stratified cluster sampling. The structure of data in the population is hierarchical, and a sample from such a population can be viewed as a multistage sample. For multistage clustered samples, the dependence among observations often comes from several levels of the hierarchy. 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. (CSA, 2011)

Multilevel analysis is a methodology for the analysis of data with complex patterns of variability, with a focus on nested sources of variability. These are used when the data structure is hierarchical with elementary units at level 1 nested in clusters at level 2, which in turn may be nested in (super) clusters at level 3, and so on. The latent variables, or random effects, are interpreted as unobserved heterogeneity at the different levels which induce dependence among all lower-level units belonging to a higher-level unit. Random intercepts represent heterogeneity between clusters in the overall response and random coefficients represent heterogeneity in the relationship between the response and explanatory variables. (Goldstein, 2003)

Logistic regression requires the assumptions of independence of the observations conditional on the explanatory variables and uncorrelated residual errors. These assumptions are not always met when analyzing nested data. But the multilevel logistic regression analyses consider 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. (Goldestein and Rasbash, 1996)

3.3.3.1 Multilevel Logistic Regression Model

3.3.3.1.1 Two-Level Model

In this study, the clustering of the data points within geographical regions offers a natural 2-level hierarchical structure of the data, i.e. children are nested within regions. Let y_{ij} be the binary outcome variable, coded '0' or '1', associated with level-one unit i nested within level two unit j. Also let p_{ij} be the probability that the response variable equals 1, and p_{ij}= Pr (y_{ij} = 1). Here, y_{ij} follows a Bernoulli distribution. Like the logistic regression the p_{ij} is modeled using the link function, logit. The two-level logistic regression model can be written as,

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right)=\beta_0 + \beta_1 x_{ij} + u_{0j} \tag{3.8}$$

Where u_{0j} is the random effect at level 2, β_0 and β_1 are coefficients

x_{ij}= Observation at i^{th} level one and j^{th} level two

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, ${\sigma_u}^2$).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,..., N) a random sample of n_j level-one units. The outcome variable is dichotomous and denoted by y_{ij} (i = 1,2,...,n_j, j = 1,2,...,N) for level-one unit i in group j.

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 ε_{ij} , that is, $y_{ij}= p_j + \varepsilon_{ij}$ the residual term is assumed to have mean zero and variance, Var (ε_{ij}) = $p_j(1-p_j)$. Then we may have the following types of models:

3.3.3.1.2 The Empty Logistic Regression Model

This is the simplest case of hierarchical two level model for a dichotomous outcome variable in which there are no explanatory variables at all. This model only contains random groups and random variation within groups.

We focus on the model that specifies the transformed probabilities $f(p_j)$ to have a normal distribution. This is expressed, for a general link function $f(p_j)$, by the formula

$$f(p_j) = \beta_0 + u_{0j} \tag{3.9}$$

where β_0 is the population average of the transformed probabilities and u_{0j} the random deviation from this average for group j. If $f(p_j)$ 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, which is expressed by

$$logit(p_j) = \beta_0 + u_{oj} \tag{3.10}$$

For the deviations u_{0j} it is assumed that they are independent random variables with a normal distribution with mean zero and variance ${\sigma_0}^2$.

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 Var(ε_{ij}) = $p_j(1-p_j)$ The probability corresponding to the average value β_0, denoted by π_0 is defined by

$$f(\pi_0) = \beta_0$$

For the logit function, the so-called logistic transformation of β_0, is defined by

$$\pi_0 = logit(\beta_0) = \frac{exp(\beta_0)}{1+exp(\beta_0)} \tag{3.11}$$

Note that due to the non-linear nature of the logit link function, there is no simple relation between the variances of the deviations u_{0j}. However, there is an approximate formula which is valid when the variances are small given by

$$var(p_j) = \left(\pi_0(1-\pi_0)\right)^2 \sigma_0{}^2 \tag{3.12}$$

3.3.3.1.3 The Random Intercept Logistic Regression Model

In the random intercept logistic regression model, the intercept is the only random effect meaning that the groups differ with respect to the average value of the response variable. But the relation between explanatory and response variables can differ between groups in more ways.

The 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. That is,

$$logit(p_{ij}) = log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_k x_{kij}$$

$$= \beta_{0j} + \sum_{h=1}^{k} \beta_h x_{hij} \tag{3.13}$$

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

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

As a result,

$$logit(p_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + u_{0j} \tag{3.14}$$

Solving for p_{ij}

$$p_{ij} = \frac{e^{\beta_0+\sum_{h=1}^{k}\beta_h x_{hij}+u_{0j}}}{1+e^{\beta_0+\sum_{h=1}^{k}\beta_h x_{hij}+u_{0j}}} \tag{3.15}$$

3.3.3.1.4 The Random Coefficient Logistic Regression Model

This is used to assess whether the slope of any of the explanatory variables has a significant variance component between the groups. Now consider a model with group-specific regressions of logit of the success probability,

$logit(p_{ij})$, on a single level-one explanatory variable X,

$$logit(p_{ij}) = log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{0j} + \beta_{1j}\, x_{1ij} \tag{3.16}$$

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

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

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

Substitution into (3.16) leads to the model

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

$$= \beta_0 + \beta_1\, x_{1ij} + u_{0j} + u_{1j}x_{1ij} \tag{3.17}$$

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_h. Thus β_1 is the average regression coefficient like β_0 is the average intercept. $\beta_0 + \beta_1 x_{1ij}$ is called the fixed part of the model and the second part, $u_{0j} + u_{1j}x_{1ij}$ is called the random part.

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. Thus, the variances and covariance of the level-two random effects $(u_{0j},\ u_{1j})$ are denoted as follows:

$$Var(u_{0j}) = \sigma_{00} = \sigma_0{}^2$$

$$Var(u_{1j}) = \sigma_{11} = \sigma_1{}^2$$

$$Var(u_{0j,}u_{1j}) = \sigma_{01}$$

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 $x_{1,}x_2, \ldots, x_k$, and consider the model where all X-variables have varying slopes and random intercept. That is

$$logit(p_{ij}) = log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} \ldots \beta_{kj}x_{kij} \tag{3.18}$$

3.3.3.2 Estimation and Testing Technique

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 (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 usage of 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)].

Parameter estimation in hierarchical generalized linear models is more complicated than the hierarchical linear models. The most frequently used kind of approximation method used are based on a first-order or second-order Taylor series expansion of the link function. In this study, the multilevel data analysis is supported by the STATA software.

3.4 Model Adequacy

Once a model is fitted to the observed values of a binary response variable, a thorough examination of the extent to which the fitted model provides an appropriate description of the observed data is a vital aspect of the modeling process. The fitted logistic regression model may be inadequate due to particular observations termed **outliers** or observations termed **influential,** values that have an undue impact on the conclusions to be drawn from the analysis and presence of complete or near complete linear dependencies among the predictor variables. Some of the statistical techniques available for the detection and treatment of outliers and influence diagnostics will be employed to examine the adequacy of a fitted model. (Goldstein, 2003)

3.4.1 Detection and Treatment of Outlier(s)

Observations that are surprisingly distant from the remaining observations in the sample are called outlying values or **outliers.** Such values may occur as a result of measurement errors, employing a faulty experimental procedure or just an extreme manifestation of natural variability. For binary data, an outlier occurs as a result of **transcription errors,** that is, when the response variable equals unity and the corresponding fitted probability is near zero or vice versa (Collet, 1991). Since fitted probabilities near zero or one will occur when the linear predictor has a large negative or a large positive value, outliers can only occur at observations that have extreme values of the explanatory variables.

Some of the statistical methods available for detecting the presence and extent of influence of outlier (s) are: unstandardized residuals $y_i - \hat{p}_i$; the Pearson or standardized residuals

$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{p}_i)^2}{\hat{p}_i(1-\hat{p}_i)}$ (The change in model deviance if the case is removed) and deviance (D) which is defined as $\pm 2\sqrt{ln\left[\frac{y_i}{\hat{p}_i}\right]^{y_i}\left[\frac{1-y_i}{1-\hat{p}_i}\right]^{1-y_i}}$. (Goldstein ,2003)

Of these methods, the study will employ the standardized residuals for identifying potential outliers. Hence, using the customary significance level of 0.01, the presence of outliers is signaled if the standardized or Pearson residuals lie outside the range of the interval (-3, +3).

Whenever outliers are identified, their effect on the results of the analysis can be passed by re-analyzing the data after omitting them. If essentially the same inferences are drawn from the data both with and without the outliers, one need not be concerned about their presence. On the other hand, if the outliers do affect model-based inferences, the decision on how to regard them (to include them, to revise the model or to omit them) should not only be made on statistical grounds but also through closer subjective and objective examination of the data and its other essential ingredients.(Goldstein ,2003)

3.4.2 Influence Diagnostics

An observation is said to be influential if its omission from the data set results in substantial changes to certain aspects of the fit of the logistic regression model. In most cases, it is more important to pay attention on outliers that are influential than those that are not. Although outliers may also be influential observations, an influential observation need not necessarily be an outlier. In particular, an influential observation that is not an outlier will occur when the observation distorts the form of the fitted model to such an extent that the observation itself has a small residual. (Goldstein, 2003)

Some of the statistical methods used in influence diagnostics are: Leverage statistic; ($\hat{h}_{ii}$ which is called the diagonal elements of the matrix $H = X\,(X\,'X)^{-1}X\,'$ with diagonal elements of $\hat{p}(1-\hat{p})$, called the "hat" matrix); DFBETA (the change in the logistic regression parameter estimates when the observation is deleted); Cook's distance C_i = $(r_i\hat{h}_{ii})/(p(1-\ \hat{h}_{ii}))$where r_i is the Pearson residual, $\hat{h}_{ii}$ the leverage of case i and p the number of parameters in the model); and the change in the Pearson statistic ($\chi^2 = \sum_{i=1}^{n}\frac{(y_i-\hat{p}_i)^2}{\hat{p}_i(1-\hat{p}_i)}$ or goodness-of-fit statistics (-2log-likelihood)).

This study will use Cook's distance to assess the overall impact of an observation on the estimated parameter vector $\hat{\boldsymbol{\beta}}$. Hence, points for which Cook's distance is greater than one will be considered as influential. Besides, the specific impact of an observation on the regression coefficients will be checked using the DFBETA statistic. Cases with DFBETA greater than unity will be considered as outlier(s) for critical variables in the model. (Goldstein ,2003)

CHAPTER FOUR: STATISTICAL DATA ANALYSIS AND RESULTS

4.1. DESCRIPTIVE ANALYSIS

Cross tabulations of the major demographic and socioeconomic background characteristics of the respondents and numbers and percentages of underweight children are presented in Table 4.1 below. The total number of children who were measured for anthropometric measurements weight-for-age was 9,526. Among these, 6,182 (64.9%) were underweight where as 3,344 (35.1%) were not underweight at the date of the survey. Out of the total number of children (9,526), the total number of female was 4,669 (49%) and the number of male was 4,857(51%).

The proportion of underweight children in Ethiopia varied from one region to the other. For example, the highest percentage (73.0%) of underweight children was observed in Tigray, followed by Benishangul Gumuz (71.1%), Somali (52.1%) and Afar (51.2%). The lowest percentage (45.5%) of underweight children was recorded for Addis Ababa. Hence, there appears to be some variation in the proportion of underweight children amongst the different regions in Ethiopia.

The proportion of underweight children, observed in Table 4.1 was also different for different age groups. For instance, higher proportions (69.4% and 69.0%) of the underweight children were in the age group 48-59 months and 36-47 months respectively. Similarly, the proportion of underweight children, observed in Table 4.1, differed by type of place of residence: urban and rural. Accordingly, higher numbers of underweight children (67.5%) resided in rural areas among 8000 children.

Table 4.1 also reveals that the proportion of underweight children varied by level of education of mothers. The highest percentage (75%) of underweight children was observed for children whose mothers have no education. When we see the level of education of partners, the highest percentage (77.8%) of underweight children was observed for fathers with no education.

Similarly, the proportion of underweight children was different depending on whether mothers were employed or unemployed. The highest percentage (73.9%) of underweight children was observed for mothers who were unemployed.

Table 4.1 also shows that the proportion of underweight children varied by households economic status. The highest percentage (81.4%) of underweight children was from poor households.

Table 4.1 shows that the proportion of underweight children varied by health and environmental characteristics. About 77.8% of the children whose family did not have access to improved source of drinking water were underweight and 77.2% of the children whose family have not toilet facilities were underweight. When we see the health status of children, 80.7% of the children who had Diarrhea in the two weeks before survey were underweight; 78.7% of the children who had fever in the two weeks before the survey were underweight and similarly 67.9% of the children who had cough in the two weeks before the survey were underweight. For the case of the other factors: sex of child, number of household members and birth order of the child, the descriptive statistics shows that the proportion of underweight children was almost similar in each category of the variable.

Table 4.1 Distribution of socioeconomic, demographic, and health and environmental characteristics

Variables	Category	No. of Underweight children	Percent	Total no. of children
Sex of child	Female	3022	64.7%	4668
	Male	3159	65.0%	4858
Age of child	0-5 month	472	48.8%	968
	6-11 months	550	57.5%	957
	12-23 months	1142	64.9%	1760
	24-35 month	1284	67.9%	1890
	36-47 months	1409	69.0%	2042
	48-59 months	1324	69.4%	1909
Birth order of the Child	0-1	1101	61.3%	1797
	2-3	1948	63.9%	3050
	4-5	1492	67.5%	2211
	6 and above	1640	66.5%	2468
Mother's Education	No educ.	4486	75.0%	5980
	Primary educ.	1491	49.9%	2990
	Secondary educ.	204	36.7%	556
Employment status of mother	Unemployed	4382	73.9%	5927
	Employed	1799	50.0%	3599
Wealth index	Poor	3990	81.4%	4903
	Medium	803	44.7%	1797
	Rich	1388	49.1%	2826
Education of partner	No educ.	3638	77.8%	4675
	Primary educ.	2079	54.4%	3823
	Secondary educ.	464	45.1%	1028
Number of household member	1-4	1535	66.0%	2324
	5-9	4144	64.3%	6443
	10 and above	502	66.2%	758
Place of residence	Rural	5397	67.5%	8000
	Urban	784	51.4%	1526
Region	Tigray	777	73.0%	1065

	Afar	618	69.4%	890
	Amhara	729	69.2%	1053
	Oromia	948	63.0%	1505
	Somali	491	65.9%	745
	Benshan-gumuz	586	71.1%	824
	SNNP	854	63.0%	1356
	Gambela	400	58.7%	682
	Harari	266	52.4%	508
	Addis Ababa	145	45.5%	319
	Dire dewa	367	63.4%	579
Source of drinking water	Not improved	3764	77.8%	4836
	Improved	2417	51.5%	4690
Had Diarrhea in the two weeks	No	4649	60.9%	7628
	Yes	1532	80.7%	1898
Had Fever in the two weeks	No	3971	59.1%	6717
	Yes	2210	78.7%	2809
Had Cough in the two weeks	No	4839	64.1%	7551
	Yes	1342	67.9%	1975
Type of toilet facility	No facilities	3900	77.2%	5049
	Have facilities	2281	50.9%	4477

Source: Own calculation based on the 2011 EDHS report

4.2 FACTORS DETERMINING WEIGHT-FOR-AGE STATUS OF UNDER-FIVE CHILDREN:-MULTIPLE LOGISTIC REGRESSION ANALYSIS

In this section, the logistic regression analysis results obtained by using stepwise inclusion of variables, overall model evaluation, statistical tests of individual predictors, goodness-of-fit statistics, and validations of predicted probabilities are presented. The model is fitted to identify the basic demographic, socio-economic, and health and environmental factors of weight-for-age status in under-five children in terms of long-run measures of health outcomes in Ethiopia.

The initial Log Likelihood Function, (-2 Log Likelihood or -2LL) is a statistical measure like total sums of squares in regression. The initial –2LL value was 12,346.500 at step 0 for underweight before any variables have been added to the model. The statistical significance of individual regression coefficients are tested using the Wald chi-square statistic. In this section, we identify the statistically significant predictor variables and determine the direction of relationship with and contribution to the dependent variable using SAS software.

Table 4.2 Test of Significance of Independent Variables Using Wald Test

Analysis of Maximum Likelihood Estimates								
Variables	DF	Estimate	S.E	Wald Chi-Square	Pr > Chi Sq	OR	95% Wald Confidence Lim.	
Intercept	1	-0.524	0.058	80.233	<.000	-	-	-
Region	10	-	-	112.861	<.000	-	-	-
Tigray	1	-0.381	0.076	25.136	<.000	0.734	0.540	0.998
Affar	1	0.581	0.084	47.318	<.000	1.921	1.395	2.644
Amhara	1	-0.075	0.076	0.972	0.324	0.997	0.731	1.360
Oromia	1	-0.136	0.063	4.686	0.030	0.938	0.697	1.260
Somali	1	0.310	0.087	12.670	0.000	1.464	1.065	2.013
Benishangul-Gumuz	1	-0.369	0.085	18.966	<.000	0.742	0.538	1.024
SNNP	1	-0.174	0.069	6.437	0.011	0.902	0.666	1.222
Gambela	1	0.238	0.088	7.243	0.007	1.362	0.985	1.883
Harari	1	0.191	0.095	4.024	0.045	1.301	0.943	1.795
Dire-Dawa	1	-0.114	0.094	1.470	0.225	0.958	0.698	1.315
Place of residence	1	-	-	8.553	0.003	-	-	-
Rural	1	0.125	0.043	8.553	0.003	1.284	1.086	1.518
Education of mothers	2	-	-	70.136	<.0001	-	-	-

No Education	1	0.689	0.121	57.183	<.000	1.993	1.572	2.525
Primary Educ.	1	0.215	0.113	3.493	0.058	1.240	.992	1.550
Source of drinking water	1	-	-	56.582	<.000	-	-	-
Not Improved	1	0.426	0.057	56.582	<.000	1.531	1.370	1.711
Type of toilet facilities	1	-	-	58.492	<.000	-	-	-
Have no facilities	1	.451	.059	58.492	<.000	1.570	1.399	1.762
Wealth index	2	-	-	182.051	<.000	-	-	-
Poor	1	0.659	0.072	167.096	<.000	1.933	1.678	2.228
Medium	1	0.407	0.042	92.799	<.000	1.240	1.145	1.531
Education of partner	2	-	-	40.7212	<.000	-	-	-
No education	1	0.390	0.096	32.352	<.000	1.477	1.224	1.783
Primary	1	0.021	0.087	9.219	0.002	1.020	0.824	1.159
Employment statusofmothers	1	-	-	147.676	<.000	-	-	-
Unemployed	1	.647	.053	147.676	<.000	1.909	1.720	2.119
Birth order of child	3	-	-	5.8762	0.118	-	-	-
1	1	-0.081	0.049	2.767	0.096	0.834	0.715	0.972
2-3	1	-0.021	0.040	0.273	0.602	0.886	0.777	1.010
4-5	1	0.002	0.045	0.001	0.973	0.906	0.788	1.042
Sex of child	1	-	-	0.0450	0.8320	-	-	-
Female	1	-0.005	0.025	0.045	0.832	0.990	0.899	1.090
Had Diarrhea in the two Weeksbefore the survey	1	-	-	59.629	<.000	-	-	-
No	1	-0.287	0.037	59.629	<.000	0.776	0.535	0.985
Had Fever in the two Weeksbefore the survey	1	-	-	104.755	<.000	-	-	-
No	1	-0.352	0.034	104.755	<.000	0.823	0.767	1.015
Had cough in the two Weeksbefore the survey	1	-	-	13.141	0.000	-	-	-
No	1	-0.128	0.035	13.141	0.000	0.775	0.675	0.889
Age of child	5	-	-	140.221	<.000	-	-	-
<6 months	1	0.598	0.066	82.838	<.000	2.444	2.037	2.932
6-11 months	1	0.258	0.067	14.973	<.000	1.740	1.446	2.095
12-23 months	1	-0.038	0.054	0.506	0.477	1.293	1.103	1.516
24-35 months	1	-0.236	0.052	20.351	<.000	1.062	0.909	1.240
36-47 months	1	-0.286	0.051	31.405	<.000	1.010	0.867	1.175

Num.of household members	2	-	-	3.034	0.219	-	-	-
1-4	1	-0.006	0.046	0.018	0.893	1.057	0.866	1.291
5-9	1	0.068	0.039	2.973	0.085	1.138	0.948	1.367

* Significant ($p<0.05$)

A negative sign in column labeled "Estimate" indicates an inverse relationship of explanatory variable with the log odds of the dependent variable. In contrast a positive coefficient in the column labeled "Estimate" indicates a positive relationship to the log odds of the dependent variable. The statistical significance of individual regression coefficients is tested using the Wald chi-square statistic. According to the results presented in Table 4.2, region, place of residence, age, education of mothers, source of drinking water, husband/partners education status, types of toilet facilities, economic status of the household (wealth index), employment status of mothers, had Diarrhea in the two weeks before survey, "fever" and "cough" in the two weeks before survey were found to be significant variables.

As can be seen in Table 4.3, at the final step, for the response variable weight-for-age status all independent variables are added to the logistic regression equation in a stepwise manner. The addition of these variables reduced the initial log likelihood value (-2 Log Likelihood) of 12,346.500 to 9,960.803.

Table 4.3: Result of Model Fit Statistics for Intercept only and Full Model

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	12348.500	10030.803
SC	12355.662	10281.461
-2 Log L	12346.500	9960.803

The difference between these two measures is the model chi-square value or likelihood ratio (2,385.278 = 12,346.500-9,960.803) that is tested for statistical significance of underweight. This test is analogous to the F-test for R^2 or change in R^2 value in multiple regressions which tests whether or not the improvement in the model associated with the additional variables is statistically significant.

Table 4.4: Test of Significance for Relationship between the Dependent and Independent Variables.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2385.6977	34	<.0001
Score	2189.2337	34	<.0001
Wald	1753.0745	34	<.0001

In Table 4.4, the model Chi-Square value of weight-for-age 2,385.6977 has a P-Value of less than 0.01. Similarly, score and Wald tests also have P-values less than 0.05. These indicate that all three tests yield similar conclusions, that is, the final model with explanatory variables was more effective than the null model. So, we conclude that there is a significant relationship between the response variable and the set of independent variables mentioned above.

The Hosmer-Lemeshow goodness of fit test was employed and the result is presented in Table 4.5. A good model fit is indicated by a non-significant chi-square value.

Table 4.5: Test of Significance of Hosmer-Lemeshow Goodness of Fit Statistics

Hosmer and Lemeshow goodness of fit test		
Chi-square	Df	Pr>Chisq
8.82	8	0.4475

The better model fit is indicated by a smaller difference in the observed and predicted classification. The Hosmer and Lemeshow Goodness of fit test tests the hypotheses:

Ho: the model is a good fit, vs.

Ha: the model is not a good fit

Since the P-value in Table 4.5 (p= 0.4475) is larger than 0.05, we do not reject the null hypothesis, and we conclude that the model is a good fit.

Table 4.6: Association of Predicted Probabilities and Observed Responses

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.0	Somers' D	0.563
Percent Discordant	21.7	Gamma	0.564
Percent Tied	0.2	Tau-a	0.257
Pairs	20669264	c	0.781

The value of the Gamma statistic is 0.564 implying that 56.40% fewer prediction errors are by using the estimated probabilities than by chance alone. The value of C statistic is 0.781. This means that for 78.1% of all possible pairs of children the model correctly assigned a higher probability to those who were underweight. The values displayed in Table 4.6 indicate that the fitted model has reasonable predictive power.

A more appealing way to interpret the regression coefficient in logistic model is using odds ratio. The odds ratio indicates the effect of each explanatory variable directly on the odds of being underweight rather than on log (odds). Estimates of odds greater than 1.0 indicate that the risk of being underweight is greater than that for the reference category.

Estimates less than 1.0 indicate that the risk of being underweight is less than that for the reference category of each variable. So, the final model presented in Table 4.2 is interpreted in terms of odds ratios as follows.

The odds of being underweight in Affar, Somali, Gambela and Harari regions are greater than one. This shows that the risk of being underweight in these regions is greater than the risk of being underweight in the reference population of Addis Ababa. The odds of being underweight in Tigray and Benishangul Gumuz are less than one. This indicates that the risk of being underweight in Tigray and Benishangul-Gumuz is less than the risk of being underweight in the reference population of Addis Ababa. In the regions Amhara, Oromia, SNNP and Dire Dawa the odds of being underweight are approximately (almost) equal to one implying that the risk of being underweight in these regions is similar to the risk being underweight in the reference population of Addis Ababa. Children in Affar are 92.1% more likely to be underweight than children in Addis Ababa controlling for other variables in the model. The odds of being underweight compared to not being underweight decreased by a factor of 0.734 when the respondent (children) is in Tigray compared to children in Addis Ababa controlling for other variables in the model.

The likelihood of being underweight was highly significant for children's aged less than five years. The risk of being underweight for children in the age groups less than 6 months and 6-11 months were 2.444 and 1.740 times the risk of being underweight for children in the age group 48-59 months respectively. Conversely, the risk of being underweight for children in the age group 24-35 months and 36-47 months is almost similar to the risk of being underweight for children in the reference age group 48-59 months.

The analysis also showed that children whose parents resided in rural areas had high risk of being underweight compared to children in urban areas. Children in rural areas were 28.4% more likely to be underweight than children in Urban areas controlling for other variables in the model.

Mother's level of education and partner's level of education are also significant factors associated with the status of weight-for-age of children. Children whose mothers had no formal education were 99% more likely to be underweight than children whose mothers had at least secondary level education. Similarly, those children whose father's had no formal education compared to children whose fathers had secondary and higher education level were 47.7% more likely to be underweight controlling for other variables in the model. Similarly, source of drinking water and types of toilet facilities are also significant factors associated with whether a child is underweight or not. Likewise, children from families with no improved drinking water sources were 53% more likely to be underweight as compared to children in the reference category. For types of toilet facilities, children of households with no toilet facilities are 57% more likely to be underweight as compared to those who have toilet facilities.

Household's wealth index also showed a statistically significant association with whether a child is underweight. The likelihood of being underweight was 93% more for children from the poor families than those from the rich families controlling for other variables in the model. The model also indicates that the odds of being underweight for children who had no diarrhea, no fever and no cough in the two weeks before the survey date were 0.776 , 0.823 and 0.775 times the odds of being underweight for children who had diarrhea ,fever and cough in the two weeks before the survey respectively. In other words, children who had no diarrhea, fever and cough are 22.4%, 17.7% and 22.5% less likely to be underweight than children who had diarrhea, fever and cough in the two weeks before the survey controlling for other variables in the model respectively. The employment status of mothers had significant effect on the status of being underweight.

The odds of being underweight for children whose mothers are unemployed was 1.909 times the odds of being underweight for children whose mothers are employed. This means, children whose mothers were unemployed were 91% more likely to be underweight than children whose mothers are employed.

4.2.1. MODEL DIAGNOSTICS

Regression model building is often an iterative and interactive process. The first fitted model for the data may be inadequate. Regression diagnostics are used to detect problems with the model and suggest improvements. A failure to detect outliers and influential cases can have severe distortion on the validity of the inferences. It would be reasonable to use diagnostics to check if the model is adequate or not. The adequacy of the fitted model was checked for possible presence and treatment of outliers and influential values. (Agresti, 2002)

The diagnostic test results for detection of outliers and influential values are presented in the Appendix A. The table shows that the normalized residuals are within the interval of -3 and 3 implying that no outliers were detected at 0.05 levels of significance. The DFBETAs for model parameters including the constant term and values of Cook's influence statistics were both less than unity. DFBETAs less than unity implies no specific impact of an observation on the coefficient of a particular predictor variable while Cook's distance less than unity showed that an observation had no overall impact on the estimated vector of regression coefficients $(\hat{\beta})$. The value of the leverage statistic less than one shows that no subject has a substantial large impact on the predicted values of model. Thus, from the above goodness of fit tests and diagnostic checking, we can say that our model is adequate.

4.3 FACTORS DETERMINING UNDERWEIGHT OF UNDER-FIVE CHILDREN:-MULTILEVEL LOGISTIC ANALYSIS

The data used in this study have a hierarchical structure. Units at one level are nested within units at the next higher level. Here, the lower level (level-1) units are the individual children, and the higher level (level-2) units are the regions that constitute the groups into which the children are clustered or nested. The nesting structure is children within regions that resulted in a set of 11 regions with a total of 9,526 children. The data used in this study consist of variables describing individuals as well as variables describing regions. Therefore, the statistical model used has to describe the data at both levels, in order to find the effect on underweight of both the individual children and the regions.

As mentioned above, multilevel models were developed to analyze hierarchically structured data. These models contain variables measured at different levels of the hierarchy. Unobserved heterogeneity is modeled by introducing random effects. Random intercepts are used to model unobserved heterogeneity in the overall response; random coefficients model unobserved heterogeneity in the effects of explanatory variables on the response variable. As one of the aims of this study was to model the heterogeneity between regions, a random intercept model was used. This allows the overall probability of to vary across regions. The chi-square test was applied to assess heterogeneity between regions mean. The test yields $X^2 = 166.507$ with d.f 10 (P=0.000). Thus, there is evidence of heterogeneity among the regions.

Children in this study were selected from different regions of Ethiopia. Thus there are two kinds of random variability in the data, that between different children in a single region, and that between different regions. The advantages of using a multilevel model include the ability to fully explore the variability at all levels of the data hierarchy, and estimation of correct standard errors in the presence of clustered data. Children from the same region would tend to be more similar compared to children chosen at random from different regions. The model takes into account the correlation structure of the data, enabling correct inferences to be made.

4.3.1. DETERMINANTS OF UNDERWEIGHT IN ETHIOPIA: INTERCEPT-ONLY MODEL WITH RANDOM EFFECT

The empty model contains no explanatory variables and it can be considered as a parametric version of assessing heterogeneity among regions with respect to underweight in under-five children. The deviance of an empty model with random effect (deviance

= 12,227.6914) is given in the Table 4.7 below. The intercept only model has AIC value 12,231.69 and BIC value 12,246.01. The variance of the random factor in the empty model is significant which indicates that there are regional differences in the weight-for-age status of under-five children.

Table 4.7: Parameter Estimate of intercept-only model with random effect
Log likelihood = -6113.846

Fixed part	Estimates	S.E	Z	P >|Z|	95% Conf. interval
Intercept	.557	.099	5.59	0.000	0.362- 0.752
Random part	Estimates	S.E	95% Conf. interval		
$Var(U_{oj})=\sigma^2_0$	0.103	.048	0.042 - 0.255		
LR test vs. logistic regression: chibar2(01) = 118.81 Prob>=chibar2 = 0.0000					
AIC= 12231.69 ,BIC = 12246.01					
Deviance =12227.6914					

4.3.2 DETERMINANTS OF UNDERWEIGHT IN ETHIOPIA: RANDOM INTERCEPT AND FIXED EFFECT MODEL

We will now extend our model to allow for regional effects on the probability of being underweight. We begin with a random intercept or variance components model that allows the overall probability of underweight to vary across regions. The results of the two-level random intercept and fixed slope model are presented in Table 4.8. STATA software has been used.

Table 4.8 Estimates for random intercept and fixed coefficient model

Log likelihood = -5004.3958, Wald chi2(18) = 1640.50, Prob > chi2 = 0.0000

Variables	Level	coefficient	S.E	Z	P>|z|	Odds ratio	95% Conf. interval	
Intercept		.140	.164	0.85	0.394	-	-	-
Place of residence	Rural	-0.256	0.084	-3.05	0.002	0.774	0.656	0.912
Place of residence	Urban(ref.)	.	.	.	.	.	.	.
Education of mothers	No education	0.648	0.119	5.45	0.000	1.912	1.515	2.414
Education of mothers	Primary Educ.	0.197	0.113	1.74	0.081	1.217	0.976	1.519
Education of mothers	Secon. educ.(ref)	.	.	.	.	.	.	.
Source of drinking water	No improved	0.431	0.056	7.62	0.000	1.538	1.377	1.719
Source of drinking water	Improved (ref)	.	.	.	.	.	.	.
Types of toilet facilities	No facilities	0.446	0.059	7.61	0.000	1.563	1.393	1.753
Types of toilet facilities	Havefacilities (ref)	.	.	.	.	.	.	.
Wealth index	Poor	0.655	0.072	9.09	0.000	1.924	1.671	2.216
Wealth index	Medium	-0.278	0.074	-3.77	0.000	0.757	0.655	0.875
Wealth index	Rich (ref.)	.	.	.	.	.	.	.
Education of partner	No educ.	0.376	0.095	3.94	0.000	1.456	1.208	1.754
Education of partner	Primary educ.	0.015	0.086	0.17	0.863	1.015	0.857	1.203
Education of partner	Secon. Educ. (ref.)	.	.	.	.	.	.	.
Employment status of	Unemployed	0.649	0.053	12.3	0.000	1.913	1.725	2.122

mothers								
Employment status of mothers	Employed	.	.	.	.	.	.	.
Had Diarrhea in the two Weeksbefore the survey	No	-0.573	0.074	-7.71	0.000	0.564	0.488	0.652
Had Diarrhea in the two Weeksbefore the survey	Yes (ref.)	.	.	.	.	.	.	.
Had fever in the two Weeksbefore the survey	No	-0.699	0.069	-10.2	0.000	0.497	0.434	0.568
Had fever in the two Weeksbefore the survey	Yes (ref.)	.	.	.	.	.	.	.
Had cough in the two Weeksbefore the survey	No	0.249	0.070	3.55	0.000	1.284	1.118	1.473
Had cough in the two Weeksbefore the survey	Yes (ref.)	.	.	.	.	.	.	.
Age of child	<6 months	-0.900	0.093	-9.71	0.000	0.406	0.339	0.487
Age of child	6-11 months	-0.559	0.094	-5.92	0.000	0.572	0.475	0.688
Age of child	12-23 months	-0.259	0.081	-3.20	0.001	0.772	0.658	0.905
Age of child	24-35 months	-0.064	0.079	-0.81	0.416	0.938	0.803	1.095
Age of child	36-47 months	-0.010	0.077	-0.13	0.893	0.989	0.850	1.152
Age of child	48-59 months (ref.)	.	.	.	.	.	.	.
Sex of child	female	0.011	0.049	0.23	0.817	1.011	0.918	1.114
Sex of child	Male (ref.)	.	.	.	.	.	.	.
Birth order of child	1	0.184	0.078	2.35	0.019	1.202	1.031	1.401
Birth order of child	2-3	0.122	0.067	1.82	0.068	1.129	0.991	1.288
Birth order of child	4-5	0.101	0.071	1.41	0.158	1.106	0.962	1.272
Birt order of child	6+	.	.	.	.	.	.	.
Number of household member	1-4	-0.054	0.102	-0.53	0.595	0.947	0.776	1.156
Number of household member	5-9	-0.128	0.093	-1.37	0.170	0.879	0.732	1.056
Number of household member	10+	.	.	.	.	.	.	.
Random-effects Parameters			Estimate		Std. Err		-	-
REG: Identity var(_cons) REG: Identity sd(_cons)			0.073 0.270		0.035 0.064		-	-

LR test vs. logistic regression: chibar2(01) = 76.04 Prob>= chibar2 = 0.0000
Deviance = 10008.7916, AIC=10048.79, BIC=10192.03

*Significance level 5%

As shown in Table 4.7 and Table 4.8, the variance between regions has decreased from 0.103 in the empty model with random intercept to 0.073 in the random intercept and fixed slopes multilevel logistic regression model and the significance of it indicates that there is a significant variation between regions in the underweight status of children.

The difference in deviance between the intercept-only model with random effect (deviance = 12,227.691) and random intercept and fixed effect model (deviance = 10,008.792) is 2,218.899 The likelihood ratio test of the null hypothesis that there is no significant difference between the two models gives LR chi2(18) = 2,218.90 (which is the difference between the deviance of random intercept only model and random intercept with random slopes model) and p-value of 0.000 implying that there is a significant difference between the two nested models. In addition, the AIC and BIC value for random intercept and fixed effect model (AIC=10,048.79, BIC=10192.03) are less than those for the intercept-only model with random effect (AIC=12,231.69, BIC = 12,246.01). This indicates that random intercept and fixed effect model is a better fit as compared to intercept-only model with random effect. In addition to this, for the test of hypothesis H_0: σ^2_u =0 which means there is no cross regional variation in under-five children, when we see the likelihood ratio test versus logistic regression, it has chibar2(01) = 76.04 Prob>= chibar2 = 0.000. Therefore the null hypothesis is rejected and there is evidence of heterogeneity or cross-regional variation in weight-for-age status of under-five children for random intercept and fixed effect model.

4.3.3. DETERMINANTS OF CHILDREN UNDERWEIGHT IN ETHIOPIA: THE 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. The variance components model which we have just specified and estimated above assumes that the only variation between regions is in their intercepts. We should allow for the possibility that the regions have different slopes. This implies that the coefficients of explanatory variables are random at level two. All variables included in the random intercept model are included in random coefficient model. Estimates of this model show that the random slope variances of all included variables except Employment status of mothers (ESM) and wealth index (WIN) are zero. This indicates that the effect of these variables is the same for each region. Results of the random coefficient model are presented in Table 4.9.

Table 4.9 Estimate of Random Coefficient Multilevel

Log likelihood = -5003.131 Wald chi2(18) = 1198.98 Prob > chi2 = 0.000

variables	Level	coefficient	S.E	Z	P>|z|	Odds ratio	95% Conf. interval	
Intercept		0.148	0.169	0.88	0.381			
Place of residence	Rural	-0.277	0.086	-3.22	0.001	0.758	0.640	0.897
Place of residence	Urban(ref.)	.	.	.	.	.	.	.
Education of mothers	No education	0.638	0.119	5.36	0.000	1.893	1.499	2.392
Education of mothers	Primary Educ.	0.184	0.113	1.62	0.105	1.202	0.962	1.501
Education of mothers	Secon. educ.(ref)	.	.	.	.	.	.	.
Source of drinking water	No improved	0.431	0.057	7.59	0.000	1.539	1.377	1.720
Source of drinking water	Improved (ref)	.	.	.	.	.	.	.
Types of toilet facilities	No facilities	0.445	0.059	7.54	0.000	1.561	1.390	1.753
Types of toilet facilities	Have facilities (ref)	.	.	.	.	.	.	.

Wealth index	Poor	0.694	0.099	7.01	0.000	2.001	1.648	2.429
Wealth index	Medium	-0.262	0.081	-3.23	0.001	0.769	0.657	0.902
Wealth index	Rich (ref.)	.	.	.	.	.	.	.
Education of partner	No educ.	0.374	0.096	3.91	0.000	1.453	1.205	1.753
Education of partner	Primary educ.	0.014	0.087	0.16	0.872	1.014	0.855	1.202
Education of partner	Secon. Educ. (ref.)	.	.	.	.	.	.	.
Employment status of mothers	Unemployed	0.665	0.060	11.07	0.000	1.944	1.728	2.187
Employment status of mothers	Employed	.	.	.	.	.	.	.
Had Diarrhea in the two Weeksbefore the survey	No	-0.567	0.074	-7.63	0.000	0.567	0.490	0.656
Had Diarrhea in the two Weeksbefore the survey	Yes (ref.)	.	.	.	.	.	.	.
Had fever in the two Weeksbefore the survey	No	-0.702	0.069	-10.2	0.000	0.496	0.433	0.567
Had fever in the two Weeksbefore the survey	Yes (ref.)	.	.	.	.	.	.	.
Had cough in the two Weeksbefore the survey	No	0.250	0.070	3.56	0.000	1.284	1.119	1.474
Had cough in the two Weeksbefore the survey	Yes (ref.)	.	.	.	.	.	.	.
Age of child	<6 months	-0.900	0.093	-9.70	0.000	0.406	0.339	0.488
Age of child	6-11 months	-0.556	0.095	-5.88	0.000	0.573	0.476	0.690
Age of child	12-23 months	-0.256	0.081	-3.16	0.002	0.774	0.660	0.907
Age of child	24-35 months	-0.064	0.079	-0.81	0.419	0.938	0.803	1.095
Age of child	36-47 months	-0.009	0.077	-0.12	0.908	0.991	0.851	1.153
Age of child	48-59 months (ref.)	.	.	.	.	.	.	.
Sex of child	female	0.025	0.059	0.26	0.837	1.021	0.938	1.154
Sex of child	Male (ref.)	.	.	.	.	.	.	.
Birth order of child	1	0.194	0.083	2.41	0.029	1.212	1.041	1.421
Birth order of child	2-3	0.142	0.083	1.89	0.076	1.132	0.993	1.289
Birth order of child	4-5	0.121	0.081	1.47	0.167	1.116	0.972	1.281
Birt order of child	6+	.	.	.	.	.	.	.

Number of household member	1-4	-0.064	0.121	-0.54	0.596	0.953	0.786	1.166
Number of household member	5-9	-0.137	0.094	-1.38	0.181	0.889	0.742	1.063
Number of household member	10+	.	.	.	.	.	.	.
Random-effects Parameters			Estimate		Std. Err			
REG: Unstructured								
var(ESM)			.0049		.0085			
var(WIN)			.0094		.0093			
var(_cons)			.0913		.0502			
cov(ESM,WIN)			.0057		.0073			
cov(ESM,_cons)			-.0004		.0211			
cov(WIN,_cons)			-.01669		.0189			
LR test vs. logistic regression: chi2(6) = 78.57 Prob > chi2 = 0.0000 AIC=10056.26, BIC= 10235.3								

Ref=Reference Category, *=significant at 5% level of significance

Table 4.10 the likelihood ratio test of models

Tests	Value	Prob>chi2
LRT (Assumption: m_1 nested in m_2)	LR chi2(18) = 2218.90	0.000
LRT (Assumption: m_1 nested in m_3)	LR chi2(23) = 2221.43	0.000
LRT (Assumption: m_2 nested in m_3)	LR chi2(5) = 2.53	0.7719

Where: m_1= Random intercept only (empty) model

m_2 = Random intercept and fixed effect model

m_3 = Random coefficient model

In Table 4.9, the value of the Wald test statistic, which is identical with Z-test for large samples for testing the significance of individual parameter, is given. The Wald test of overall model fit with Wald chi2(18) = 1198.98 and 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 indicates that there is evidence that the logistic regression model with no intercept estimated 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 fixed intercept is 0.148 and the odds of the probability of child being underweight when all level one covariates are zero in region j is given by $0.148+\hat{u}_j$ where $\hat{u}_j$ is a random intercept with a variance of .0913 (indicated in the table as var(cons)) which is the between-regions variance and standard error 0.050. In the absence of level-one covariates the effect of each region on whether a child is underweight as compared to the weight of the average child measured with odds depends on the sign of the random intercept, $\hat{u}_j$. When $\hat{u}_j$ is positive the odds of an under-five child being underweight is higher than the average and when $\hat{u}_j$ is negative the odds of a child being underweight is less than the average. The individual region slopes of employment status of mothers (ESM) and wealth index (WIN) vary with a variance of 0.005 and 0.009, respectively. The negative covariance estimate of -0.00031 (standard error 0.0211) between intercept and slopes of employment status of mothers and the negative covariance estimate of -0.0167 (standard error 0.0189) between intercept and slopes of wealth index suggest that regions with a high intercept (above-average) tend to have a flatter-than-average slope. Generally, interpretation of significant covariance terms can be easily made in terms of the correlation coefficients between random intercept and random slopes. Positive covariance/correlation between intercept and slopes implies that regions with higher intercepts tend to have on average higher slopes on the corresponding predictors.

The intercept-slope correlation between intercept, slope of employment status of mothers and wealth index is estimated as follows:

$\rho_{02}= \frac{-0.00031}{\sqrt{0.00499*0.09131}}$ = -0.0145 (correlation between intercept and slope of Employment status of mothers)

$\rho_{03}= \frac{-0.0167}{\sqrt{0.0094*0.09131}}$ = -0.5700 (correlation between intercept and slope of wealth index)

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. This value indicates that children who live in regions with higher proportion of employed mothers and high wealth index are less likely to be underweight than children who live in regions with higher proportion of unemployed mothers and low wealth index.

In order to see whether the inclusion of level one covariates with employment of mothers and wealth index varied across regions significantly improved the random intercept only model, log- likelihood ratio test, AIC and BIC values were used. The likelihood ratio test of no significant difference between the two models gives LR chi2(18) = 2,221.43 (which is the difference between the deviance of random intercept only model and random coefficient model) and p-value of 0.000 which implies that there is a significant difference between the two nested models. The AIC and BIC values of the random intercept only model are greater than the corresponding AIC and BIC values of the random coefficient model indicating that the random coefficient model is a better fit than the random intercept only model. Similarly the likelihood ratio test of no significant difference between the random intercept and fixed effect model and random coefficient models resulted in LR chi2(5) = 2.53 (which is the difference between the deviance of random intercept only model and random coefficient model) with p= 0.7719 which implies that there is no significant difference between the two models. Finally using the model comparison methods such as deviance, AIC and BIC the random intercept and fixed effect model provided a good fit to the data.

4.3.4. GOODNESS OF FIT TEST

An overall evaluation of the multilevel logistic regression model was assessed using the deviance and likelihood ratio test. The likelihood ratio test of no significant difference between the given models can be done by using the STATA output. The significance of this likelihood ratio test indicates that the model is a good fit or is a better model than the smaller model. Similarly, the overall model evaluation was assessed using AIC and BIC. Based on the results in Table 4.10 (random coefficient model), the likelihood ratio test is significant and the values of AIC and BIC are used for model comparison between random intercept, random intercept and fixed effect and random coefficient models.

CHAPTER FIVE: DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

In this chapter the main findings in this study are discussed, conclusions drawn and recommendations made.

5.1. DISCUSSION

This study was an attempt to identify the determinants of weight-for-age status (whether a child is underweight or not) among under-five children in Ethiopia based on Ethiopian Demographic and Health Survey (EDHS 2011) data. Accordingly descriptive analysis, multiple logistic regression and multilevel logistic regression techniques were used.

The study included the most important predictor variables that were categorized under socio-economic, demographic, health and environmental characteristics. Concerning the regional disparity children's in Affar, Somalia, Gambela and Harari are at a higher risk of underweight than children who live in Addis Ababa.

In our study, the difference in the prevalence of underweight between male and females was almost insignificant. Gender (being female or male) had no effect on whether a child is underweight or not in the present study. This finding contradicts the findings in a number of studies in Africa.The studies indicated that rates of malnutrition among boys are consistently higher than among girls (Svedberg, 1990; MoFED, 2002a; Glewwe *et al.*, 2002; Sahn and Stifel, 2002; Christiaensen and Alderman, 2004). Sahn and Stifel (2002).

The prevalence of underweight in our study was higher in rural than urban areas.

The present study indicated that age of child is also one of the important determinant factors of weight-for-age status of under-five children in Ethiopia. The prevalence of underweight was higher among children aged <6 months and 6-11 months than in the other age groups. This finding supports the findings of previous studies conducted by Md.Israt and Sekander, 2006 who found out that children are highly underweight in this age interval, the reason could be that children may be born with low weight.

In this study, prevalence of underweight and education of mothers and fathers were positively related. A larger proportion of children whose mothers had no education or primary education were underweight compared to children whose mothers have secondary and higher level of education. This finding seems to be consistent with studies by Md.Israt and Sekander (2006) and various studies concluded that parental education, especially mothers' education, is a key element in improving children's nutritional status (Moen, 1993; Christiaensen and Alderman, 2004). Glewwe (1999) investigated the mechanisms through which education (schooling) results in better child nutrition or health.

In addition, the study revealed that under-five children from poor households were at a higher risk of being underweight than children from rich households. This finding is consistent with other studies (Stalin p. et al., 2013).They indicated that better off households had better access to food and higher incomes than poor households, allowing them a quality diet, better access to medical care and more money to spend on essential non-food items such as schooling, clothing and hygiene products. In addition to this, Glewwe *et al.* (2002) observed that, due to rapid economic growth in Vietnam since 1986, a dramatic decline in poverty and child underweight was achieved. Based on their empirical investigation, they concluded that increase in household income, had a positive impact on child nutrition in Vietnam during the 1990s. They also noted that, over time, child underweight declined within each quintile even after adjustment for change in income was made, which suggested that there were other factors, in addition to income growth, which led to improvements in child nutrition.

The findings of this study showed that children who had diarrhea, fever and cough two weeks before the date of the survey were significantly underweight than those who had not. The prevalence of underweight and diarrhea, fever and cough were related.

This study examined individual and community-level (regional) factors as significant determinants of weight-for-age status in under-five children. It confirms the importance of regional variations with respect to childhood weight-for-age status.

Using multilevel logistic regression method allowed to study variations in childhood weight-for-age status among geographical regions. The model suggested that status of children being underweight differs among regions, although the variations among different communities/regions with respect to the odds of having childhood underweight were found to originate mainly from variations in individual-level factors.

5.2 CONCLUSION

The study identified some socio-economic, demographic and health and environmental variables as determinants of weight-for-age status of under-five children in Ethiopia. Age of child, Source of drinking water, Types of toilet facilities, Mother's educational level, Employment status of mothers, wealth index, place of residence, had diarrhea, fever and cough in the two weeks before survey, educational level of partner and geographical region were found to be important determinants of weight for age status of under-five children in Ethiopia. Childhood malnutrition remains a major public health problem in Ethiopia.

Although there is region-wise disparity in children's health/nutritional status, it is observed that children living in rural parts of the country were at higher risk of being underweight. The result also suggested that children from poor families were more likely to be underweight than children from rich families in Ethiopia. Age of child, mother's educational level, partner's educational level, diarrhea, fever, cough, source of drinking water, employment status of mothers, wealth index and place of residence are significant determinants of variation in childhood weight-for-age status (whether a child is underweight or not) of under-five children among regions.

5.3 RECOMMEDATIONS

The findings from this study lead to the following recommendations:

- Create awareness about child feeding practice in scientific way.
- Organize adult literacy programs particularly for women with special focus on children health and nutrition.
- Any intervention by governmental and non-governmental organizations that aims at improving under-five children's nutritional status should consider regions with a high rate of childhood underweight so as to avert under-coverage of the regions that deserve it.
- Nutritional intervention programmes like the therapeutic feeding and nursery school feeding programmes should be established and be directed specifically towards higher risk groups such as children born to non-educated or less-educated mothers. This will at least lessen the degree of linear growth retardation in those who are already experiencing childhood underweight.

REFERENCES

Agresti, A. (2002). An Introduction to Categorical Data Analysis. 2nd Edition John Wiley and Sons Inc., New York.

Alderman, H., Behrman, J., and Hoddinott, J. (2004). Nutrition, malnutrition and economic growth in Health and economic growth: findings and policy implications. López-Casasnovas,G., Rivera, B. andCurrais, L., Cambridge,MA: MIT Press

Alderman, H., Hoddinott, J. and Kinsey B. (2002). Long term consequences of early childhood malnutrition. Mimeo, Department of Economics, Dalhousie University, Halifax, Canada

Ambrus JL Sr. and Ambrus JL Jr. (2004). Nutrition and infectious diseases in developing countries and problems of acquired immunodeficiency syndrome. Exp Biol Med (Maywood); **229:**464–72.

Baradaran, R. and Hasani, K. (2006).Malnutrition in children. Tehran: Iranian Pediatrics Association Publisher; 883-886.

Basu, K. and Foster, J. (1998). On measuring literacy, Economic Journal,108:1733-49

Bender, D.A.(1997). Introduction to Nutrition and Metabolism (2nd Ed.). London: Taylor and Francis.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. J. Am. Statist. Assoc.

Brown,J.E. (2002). Nutrition through the life cycle, Belmont, CA: Wadsworth and Thomson Learning

Christiaensen, L. and Alderman, H. (2004). Child malnutrition in Ethiopia: can maternal knowledge augment the role of income: Economic Development and Cultural Change, **52(2)**:287-312

Collet, D. (1991). Modeling Binary Data. Chapman & Hall, London.

Copas, J.B. (1988). Binary Regression Models for Contaminated Data. Journal of Royal Statistical Association, B 50 (2).

CSA (Central Statistical Authority) (2000). Demographic and health survey, Addis Ababa

De Onis M, Monteiro C, Akre, J. and G., Clugston (2003).The worldwide magnitude of protein-energy malnutrition: an overview from the WHO Global Database on Child Growth. In: http://www.who.int/whosis/cgrowth/bulletin.htm.

De Onis, M., Blössner M. (2003). The World Health Organization Global Database on Child Growth and Malnutrition: methodology and applications. Int J Epidemiol; **32**:518–526.

Degefe, B. and Nega, B. (2000). Annual report on the Ethiopian economy, **I**, Addis Ababa: Ethiopian Economic Association

Desai, (1995). When are children from large families disadvantaged? Evidence from cross-national Analyses, Population Studies **49 (2):** 195-210

EDHS (2005). Ethiopian Demographic and Health Survey. Addis Ababa

EDHS (2011). Ethiopian Demographic and Health Survey. Addis Ababa

Ehtisham Ahmad, Salman Khalil and Zulfia khan (2011). Nutritional Status in Children (1-5 Yrs)-A Rural Study. Indian Journal of Community Health. **23(2)**: 84 – 86.

Esrey S., Habicht, J-P. and Casella, G. (1992).The complementary effect of latrines and increased water usage on the growth of infants in rural Lesotho: American Journal of Clinical Nutrition, **135 (6)**:659-666

Esrey, S. (1996). Water, waste and well-being: a multi-country study, American Journal of Clinical Nutrition, **143 (6)**: 608-623

Gibson,J. (2001). Literacy and intra-household externalities, World Development, **29**: 155-166

Girma, W. and Genebo, T. (2002). Determinants of nutritional status of women and children in Ethiopia. Calverton, Maryland: ORC Macro

Glewwe, P. (1999). Why does mother's schooling raise child health in developing countries? Evidence from Morocco, Journal of Human Resources, **34(1)**: 124-159

Glewwe, P. Koch, S. and Nguyen, B.L. (2002). Child nutrition, economic growth and the provision of healthcare services in Vietnam, February,Washington:World Bank

Goldstein, H. (1995). Multilevel Statistical Models. London: Edward Arnold/New York: Halsted.

Goldstein, H. (2003). Multilevel Statistical Models. (3rd Ed.). London: Arnold; New York: Oxford University Press Inc.

Goldstein, H. and J. Rasbash., (1996). Improved approximations for multilevel models with binary responses. J. Roy. Statist. Soc. A.

Goldstein, H., (1991). Nonlinear multilevel models with an application to discrete response data. Biometrika, **78**,45–51.

Hosmer, W.D. and S. Lemeshow (2000). Applied Logistic Regression. 2nd Ed., John Wiley and Sons, New York.

Macfarlane,S.B. (1996). The relationship between anthropometric indices derived from the CDC/ WHO international reference', Annals of Human Biology: **23 (6)**: 421-30

Maddala, G.S. (1997). Introduction to Econometrics.2nd Ed. Prentice-Hall, Inc., USA.

Md.Israt Rayhan and M.Sekander (2006). Factors causing Malnutrition among under five children in Bangladish.Pakistan journal of nutrition **5(6)**:558-562

Moen, A.A. (1993). The impact of child and maternal survival programmes on socio-economic development, Center for Economic Research on Africa, School of Business, Montclair State University and Upper Montclair, New Jersey

MoFED (Ministry of Finance and Economic Development) (2002a). Development and poverty profile of Ethiopia, Welfare Monitoring Unit, March, Addis Ababa

MoFED (Ministry of Finance and Economic Development) (2002b). Ethiopia: sustainable development and poverty reduction programme, Welfare Monitoring Unit, July, Addis Ababa

MoH (Ministry of Health) (2002). Health Sector Development Phase II, Addis Ababa

Murray CJL and AD Lopez (1996). The global burden of diseases: A comparative assessment to mortality, disability from disease, injury and risk factors in 1990 projected to 2020. The Global Burden of Diseases and Injury Series, vol 1.Harvard School of Public Health, Cambridge. M.A.

National Institute of Population Studies (1992). Pakistan Demographic and Health Survey 1990/91. NIPS Islamabad Pakistan, p. 149–63.

Pelletier, D.L., Frongillo, Jr., E.A. and Habicht, J.P. (1995). The effect of malnutrition on child mortality in developing countries. Bulletin of the World Health Organization **73 (4)**:443–448.

Pelletier, DL.and EA. Frongillo (2002). Changes in child survival are strongly associated with changes in malnutrition in developing countries. Washington D.C.: Food and Nutrition Technical Assistance (FANTA) Project, Academy for Educational Development.

Phengxay, M., Ali, M., Yagyu, F., Soulivanh, P., Kuroiwa, C., Ushijima, H. (2007). Risk factors for protein energy malnutrition in children under 5 years: study from Luangprabang province, Laos. Pediatr **49(2)**: 260-265.

Rapaport, R., Bowlby, DA. (2004). Clinical aspects of growth and growth disorder. In:Pescovitsozz OH, Eugster EA. Pediatric endocrinology: Mechanisms, Manifestations and Management. Philadelphia: Lippincott Williams & Wilkins; p.172-187.

Roberts, J. (2003). Poverty reduction outcomes in educationand health:public expenditure and aid, London: Overseas Development Institute, Working Paper 210

Rossouw, K. (1996). Nutrition and anthropometric measurement in Africa, **29**

Rowland, MG., Rowland, SG and TJ. Cole (1988). The impact of nutrition on the growth of children from 0-2 years in age in an urban West African community. *Am. J. Clin. Nutr.*; **47**:134-148.

Sahn, D. E. and Stifel D. C. (2002). Parental preferences for nutrition of boys and girls: evidence from Africa, Journal of Development Studies, **39 (1)**: 21-45

Schaible UE, Kaufmann SH. Malnutrition and Infection: Complex Mechanisms and Global Impacts PLoS Med 2007; **4(5)**:e115.

Schroeder DG and KH Brown (1994). Nutritional status as a predictor of child survival: Summarizing the association and quantifying its global impact. *Bull. WHO,* **72**:569-79

Stalin P., Joy Bazroy, Dinesh Dimri, Zile Singh and Senthilvel V. (2013). Prevalence of Underweight and its Risk Factors among Under Five Children in a Rural Area of Kancheepuram District in Tamil Nadu, India (Issue 6 (Jan.- Feb.), **3**, 71-74)

State of the world's children (2007). Available: http://www.unicef.org/sowc07 .

Svedberg, P. (1990). Undernutrition in Sub-Saharan Africa: is there a gender bias? Journal of Development Studies, **26**: 469-486

UNDP (2002). Human Development Report, New York: Oxford University Press

UNICEF (2007). Progress for Children: A World Fit for Children Statistical Review (6).

UNICEF. The State of world Children. New York: Oxford University Press; 1998. p.10.

UNICEF.Analysis of the number of underweight children in the developing world. UNICEF Progress for Children. A report card on nutrition. Last updated May 2006. www.childinfo.org

United Nations Sub-Committee on Nutrition 5th Report on the World Nutrition Situation: Nutrition for Improved Outcomes. March 2004.

WHO Working Group (1986). Use and interpretation of anthropometric indicators of nutritional status. Bull World Health Organ; **64**:924–41.

WHO.Nutrition:Challenges.2005.Available at http://www.who.int/nutrition/challenges.

Woodward B. (1998). Protein, calories, and immune defenses. Nutr Rev; **56**:S84–S92.

World Bank (2004). World Development Report, New York: Oxford University Press

World Bank and MoH, (2005). National Strategy for Infant and Young Children, Addis Ababa

APPENDEX A

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
	Step	2385.698	34	.000
Step 1	Block	2385.698	34	.000
	Model	2385.698	34	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	9960.803[a]	.222	.305

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.82	8	.4475

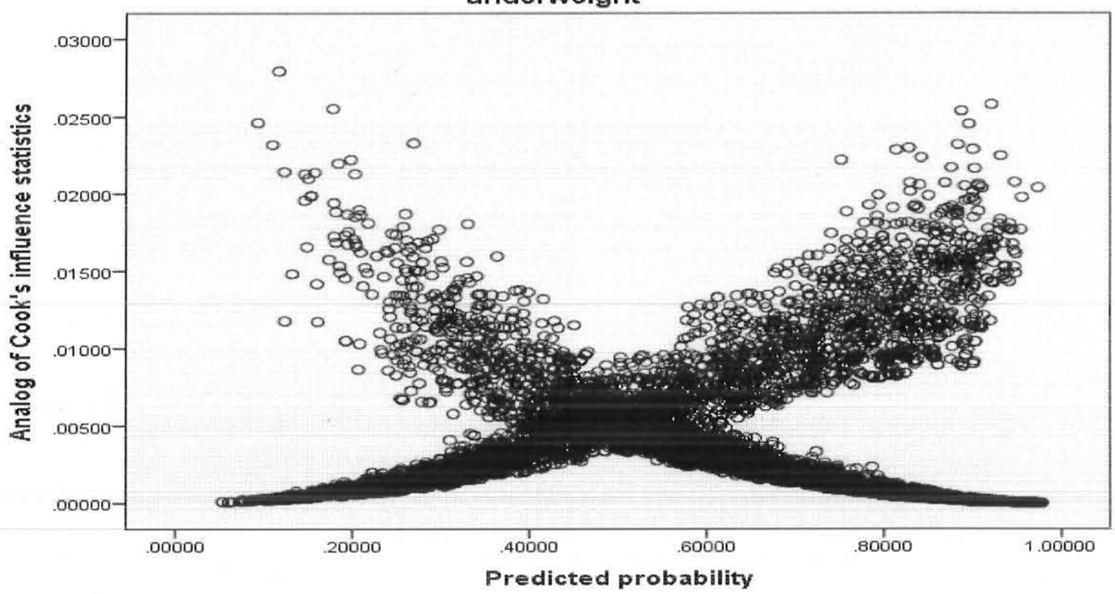
Result of diagnostic tests for outliers and influential value for underweight

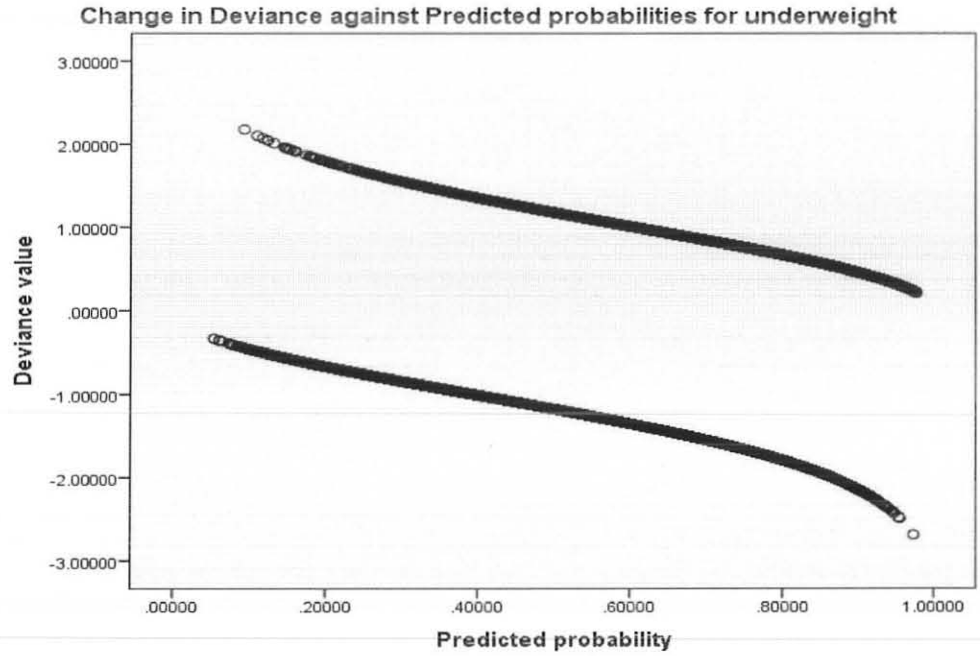
Descriptive Statistics

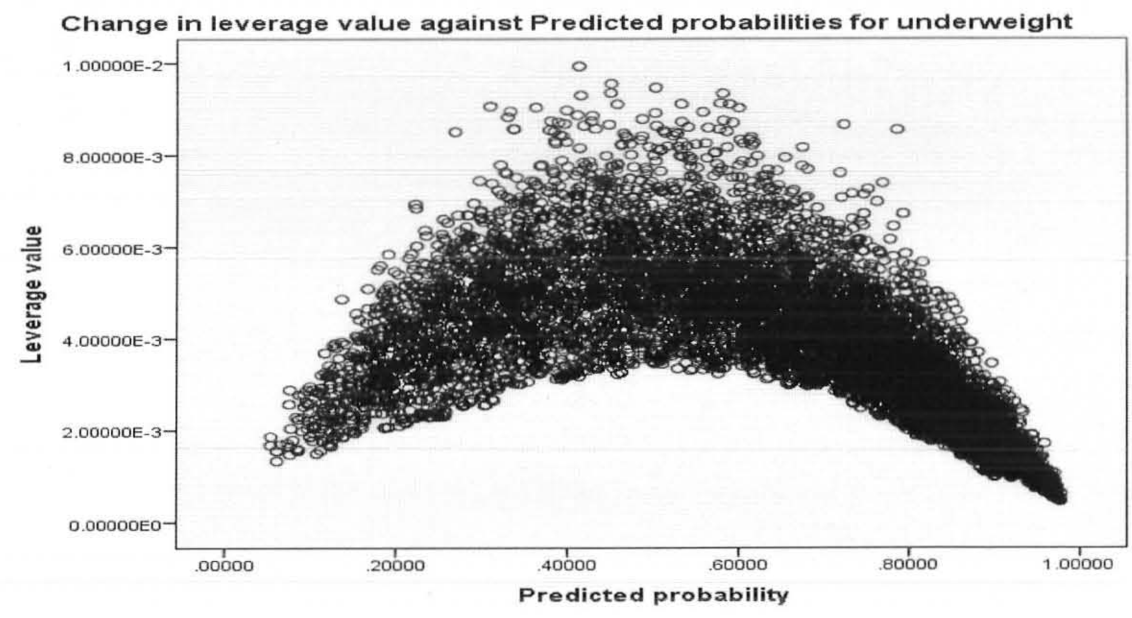
	N	Minimum	Maximum
DFBETA for constant	9525	-.01493	.01957
DFBETA for REG(1)	9525	-.01123	.01301
DFBETA for REG(2)	9525	-.01129	.01455
DFBETA for REG(3)	9525	-.01142	.01333
DFBETA for REG(4)	9525	-.01148	.01230
DFBETA for REG(5)	9525	-.01144	.01370
DFBETA for REG(6)	9525	-.01171	.01281

DFBETA for REG(7)	9525	-.01131	.01183
DFBETA for REG(8)	9525	-.01218	.01272
DFBETA for REG(9)	9525	-.01154	.01255
DFBETA for REG(10)	9525	-.01141	.01335
DFBETA for PR(1)	9525	-.00457	.00671
DFBETA for EDM(1)	9525	-.00917	.01195
DFBETA for EDM(2)	9525	-.00901	.01057
DFBETA for SDW(1)	9525	-.00214	.00260
DFBETA for TTF(1)	9525	-.00274	.00268
DFBETA for WIN(1)	9525	-.00320	.00450
DFBETA for WIN(2)	9525	-.00277	.00368
DFBETA for EDP(1)	9525	-.00593	.00766
DFBETA for EDP(2)	9525	-.00532	.00606
DFBETA for ESM(1)	9525	-.00191	.00207
DFBETA for BOC(1)	9525	-.00405	.00319
DFBETA for BOC(2)	9525	-.00278	.00246
DFBETA for BOC(3)	9525	-.00285	.00252
DFBETA for SC(1)	9525	-.00131	.00126
DFBETA for HDW(1)	9525	-.00357	.00485
DFBETA for HFW(1)	9525	-.00356	.00440
DFBETA for HCW(1)	9525	-.00370	.00405
DFBETA for AC(1)	9525	-.00481	.00535
DFBETA for AC(2)	9525	-.00523	.00519
DFBETA for AC(3)	9525	-.00336	.00377
DFBETA for AC(4)	9525	-.00325	.00340
DFBETA for AC(5)	9525	-.00286	.00314
DFBETA for NHM(1)	9525	-.00687	.00759
DFBETA for NHM(2)	9525	-.00694	.00749
Valid N (listwise)	9525		

Change in analog of cook's influence statistics against Predicted probabilities for underweight




Change in Deviance against Predicted probabilities for underweight
3.00000
2.00000
1.00000
.00000
-1.00000
-2.00000
-3.00000
Deviance value
.00000
.20000
.40000
.60000
.80000
1.00000
Predicted probability


Change in leverage value against Predicted probabilities for underweight
1.00000E-2
8.00000E-3
6.00000E-3
4.00000E-3
2.00000E-3
0.00000E0
Leverage value
.00000
.20000
.40000
.60000
.80000
1.00000
Predicted probability

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 Getachew Tadesse

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 (Assistant Prof.)

Signature____________________

Date 26/06/2013