MULTILEVEL LOGISTIC MODELLING OF UNDER-FIVE CHILD MORTALITY VARIATIONS AMONG REGIONAL STATES OF ETHIOPIA

THESIS

BY

ASHENAFI SENBETA BEDANE

HAWASSA UNIVERSITY

HAWASSA, ETHIOPIA

JUNE, 2012

ABSTRACT

Objective: *The main aim of this study is to investigate under-five child mortality variations among regional states of Ethiopia.*

Data: *This study is conducted based on Demographic and Health Survey (DHS) 2011 data, collected for 10,156 children under-five years of age in Ethiopia.*

Methods: *In this study, single level and multilevel binary logistic regression models is adopted for the analysis.*

Results and conclusions: *Based on the model adequacy tests the random intercept binary logistic regression model is found to be best fitting to the data. The variance of the random component model related to the intercept term is statistically significant, implying the presence of under-five child mortality variations among regional states of the country and it is accounted by the random intercept term. The major significant factors affected under-five child mortality are: mother's education level, birth index, child size at birth, mother's age at birth, type of birth and breastfeeding status. It also revealed that there is a contribution of those major factors to under-five child mortality variations among regional states. However, those factors significantly affecting under-five child mortality is explicitly did not show significant effects on variations of under-five child mortality across regional states. The study recommend all regional states to makes remedial measures on public health policy, design strategy to improve facilities and aptitudes of stakeholder living in their region toward those major factors affecting under-five child mortality and contributing to its variations among regional states to reduce under-five child mortality in the country.*

1. Introduction

1.1. Background

Improvements in child survival have been one of the major targets of development programs of Ethiopia during the past three decades. According to Ethiopian 2010 MDGs Reports trends and prospective for meeting MDG by 2015, sixteen out of every hundred children born in Ethiopia will not live beyond their fifth birth day. Five will not even live the first month of life. Every year, approximately, 472,000 children under five years of age die in Ethiopia (MoFED, 2010).

In 2000, the Ethiopia Demographic and Health Survey (EDHS) estimated the under-five mortality rate (U5MR) at 166 per 1000 live births. The Ethiopian Ministry of Health (MOH) estimates that the U5MR for 2002-03 was 140. In 2000, the rate of 166 placed Ethiopia at 21st in the world for under-five mortality. Ethiopia neonatal mortality rate was relatively even higher and it was the fifth-highest in the world. According to EDHS 2005, the rate of mortality in Ethiopia examined by comparing data from 2005 EDHS and 2000 EDHS, infant and under-five mortality rates, obtained for the five years preceding the two surveys, confirm a declining trend in mortality. Under-five mortality declined from 166 deaths per 1,000 live births in the 2000 survey to 123, while infant mortality declined from 97 deaths per 1,000 live births in the 2000 survey to 77 (CSA, 2005). Even the child mortality has declined in Ethiopia still it is high child death at national level as compared to developed countries. Ethiopian MDG 4 would further reduction of child and maternal death.

The regional difference in child mortality may be due to differences in socio-economic composition (Kandala et al., 2007), health-seeking behavior regarding child immunizations (Antai, 2009) and maternal and child health care utilization (Antai et al., 2009). Indeed, the incorporation of community-level factors in the analysis of child mortality provides an opportunity to identify the health risks associated with particular social structures and community ecologies, which is a key policy tool for the development of public health interventions (Pickett and Pearl, 2001; Stephenson et al., 2006).

The number of under-five deaths in worldwide has declined from more than 12 million in 1990 to 7.6 million in 2010. Nearly 21,000 children under-five died every day in 2010 which was about 12,000 less than in 1990. Since 1990, the global under-five mortality rate has dropped 35 percent from 88 deaths per 1,000 live births in 1990 to 57 in 2010. The Northern Africa, Eastern Asia, Latin America and the Caribbean, South-eastern Asia, Western Asia and the developed regions have reduced their under- five mortality. The rate of under-five mortality was reduced in the year over 2000 to 2010, but remains insufficient to reach MDG 4, particularly in Sub-Saharan Africa, Oceania, Caucasus and Central Asia, and Southern Asia (United Nation, 2011).

The highest rates of child mortality have been still in Sub-Saharan Africa where 1 in 8 children died before age five, which is more than the average for developed regions and Southern Asia. Under-five mortality rates have fallen elsewhere and the disparity between these two regions and the rest of the world has grown. Under-five deaths are increasingly concentrated in Sub-Saharan Africa and Southern Asia, while the share of the rest of the world dropped in 2010 (United Nation, 2011).

A consistent series of estimates of under-five mortality rate provided historical trends, during the period of 1950–2000 for both developed and developing countries. On an average about 15% of newborn children in Africa are expected to die before reaching their fifth birthday (Omar et al., 2000). The progress of infant and child mortality in Sub-Saharan Africa remained as a major health problem, and the progress made during the past four decade has been unevenly distributed (Garenne and Gakusi, 2006).

The neonatal, post-neonatal, infant and child mortality pattern are higher for mothers who are under 20 years of age. Infant and child mortality levels are lower for children whose mother's age is between 20 up to 29. Neonatal mortality of the children whose mothers aged is below 20 years at the time of the child's birth, is higher than the children whose mothers are in the age range 20-29 years at the time of giving birth. Short birth intervals were significantly reduced infant probability of survival. The researchers used cross classification percentage distribution and logistic regression model (Mondal et al., 2009).

An investigation on historical and modern third world countries have shown that children who are exclusively breast-fed survive longer and are healthier than artificially fed children in direct (Lindstrom et. al., 1999). And also the breastfeeding practices have significantly lower risk among neonatal, post-neonatal and child mortality levels as compared to children never breastfed (Mondal et al., 2009).

In Ethiopia a retrospective birth history data from a national survey used proportion hazard regression model (Lindstrom and Gebre-Egziabher, 2001) found a significantly higher risk of a conception in the months, following the death of an index child, even after controlling for breastfeeding status.

Maternal education has been identified as one of the most important socio-economic determinants of infant and child mortality. The study revealed a significant association between mother's education and infant and child mortality. The researchers used statistical test of independence based on Chi-square (Mahfouz et al., 2009). There has been considerable decline in infant mortality as mother's educational attainment increases. However, there exists a marked differential among the regions. Infants born to women with no education are almost more likely to die before age one than infants born to women with primary and higher education and the researcher adopted multivariate logistic regression model in Ghana (Goro, 2007).

Improvements in child survival have been one of the major targets of development programs during the past three decades. A century later, out of the 187 countries, only nineteen countries- all in Africa- had an infant mortality rate of above ten percent. Ethiopia, through the progressive implementation of the Health Sector Development Program in the last seven years, has made great strides to improve maternal and child survival. The reduction of infant and child mortality indirectly helps in reducing fertility by decreasing the desired number of children to be born due to increased probability of survival of a child. Under-five mortality is significantly influenced by breastfeeding status, ownership of toilet facilities, the level of education of the mother, residential area and place of delivery of the child; rural mothers and children are particularly at disadvantage with regards to basic health and socio-economic services based on logistic regression analysis and Cox regression (Zeleke, 2009). Birth interval with previous child and mother standard of living index are the vital factor associated with child mortality. The cross-tabulation analysis (Kumar and

Gemechis, 2010) shows that birth interval with previous child and mother standard of living index is the vital factor associated with child mortality.

Regional disparities in under-five child mortalities are associated with factors at the community level that distinguish these regions from each other. The availability of services and social amenities in communities, or the lack infrastructure, may positively or negatively influence the health of the residents of communities. Some of these factors include differences in community-level development, population density, prevalence of poverty, and availability of maternal and child health care services. These are often interrelated aspects of the regional environment that are important for child health and well-being, and may also be relevant in exacerbating or mitigating inequities in resources and population health outcomes across regions (Siddiqi et al., 2007).

The most recent studies related to child mortality in the regions within geographically diverse ecology and socioeconomic circumstances may have different disease exposures and child health outcomes. Antai (2011) tried to assess variations in the risks of death in children under age 5 across regions of Nigeria and determined characteristics at the individual and community levels that explained possible variations among regions. The researcher applied multilevel Cox proportional hazards analysis using a nationally representative sample of 6,029 children from 2,735 mothers aged 15-49 years and nested within 365 communities from the 2003 Nigeria Demographic and Health Survey. Hazard ratios (HR) with 95% confidence intervals (CI) were used to express measures of association among the characteristics. Variance partition coefficients and Wald statistic were used to express measures of variation. From the results, the researcher suggested the need to differentially focus on community-level interventions aimed at increasing maternal and child health care utilization and improving the socioeconomic position of mothers, especially in disadvantaged regions such as the South (Niger Delta) region (Antai, 2011).

Statement of the Problem

Under-five children mortality in Ethiopia is one of the highest in the world and it is one of the challenging problems that the country needs to address. Even in an average year, the education, health and economic situation for millions of Ethiopian under-five children can only be described as a crisis. In Ethiopia factors such as, low level of mother's education, unsafe drinking water and sanitation, low family income, birth interval, short to breast feeding time, lack of place of birth delivery and periodic famine continue to put children at risk.

This study has been highly motivated to investigate the major determinants of under-five child mortality and hence, it is aimed to address the following questions:

- What are the factors have significant impacts on under-five child mortality among variables considered in Ethiopia?
- Are there significant variations of under-five child mortality across the regional states of Ethiopia?
- What factors have made significant contribution to the variation of under-five child mortality among regional states of Ethiopia?

2. Data and Methodology

2.1. Descriptions of Study Area and Population

Ethiopia is officially known as the Federal Democratic Republic of Ethiopia, is a landlocked country located in the Horn of Africa. It is the second-most populous nation in Africa, with over 82 million populations (CSA, 2012) and the tenth largest by area, occupying 1,100,000 km^2. Ethiopia is bordered by Eritrea to the North, Djibouti and Somalia to the East Sudan and South Sudan to the West, and Kenya to the South. Ethiopia has eleven geographic or administrative regions: nine regional states (Tigray, Affar, Amhara, Oromia, Somali, Benishangul-Gumuz, SNNPR, Gambela and Harari) and two city administrations (Addis Ababa and Dire Dawa that are considered as region) with capital city of Addis Ababa.

Administratively, each of the 11 geographic regions in Ethiopia is divided into zones and each zone is divided into lower administrative units called *woredas*. Each *woreda* is then further subdivided into the lowest administrative unit, called a *kebele*.

2.2. Data

The data used for this study is 2011 Ethiopia Demographic and Health Survey (2011 EDHS). The survey was conducted under the guidance of the Ministry of Health by the Central Statistical Authority from 27, December 2010 through June 2011 with a nationally representative sample of nearly 18,500 households. But in this study, the data from Somali region was excluded from this study, because in the Somali region, in 18 of the 65 selected EAs listed households were not interviewed for various reasons, such as drought and security problems, and 10 of the 65 selected EAs, were not listed due to security reasons. Therefore, the data for Somali may not be totally representative of the region as a whole (CSA, 2011).

The sample for the 2011 EDHS designed to provide population and health indicators at the national and regional levels. The sample design allowed for specific indicators, such as contraceptive use, to be calculated for each of Ethiopia's eleven geographic/administrative regions: nine regional states and two city administrations. The sampling frame used for the 2011 EDHS was the Population and Housing Census conducted by the Central Statistical Authority (CSA) in 2007. During the 2007 PHC, each of the *kebeles* was subdivided into convenient areas called census enumeration areas (EAs). The 2011 EDHS sample was selected using a stratified, two-stage cluster design and EAs were the sampling units for the first stage. The 2011 EDHS sample included 624 EAs, 187 in urban areas and 437 in rural areas (CSA, 2011).

Households comprised the second stage of sampling. A complete listing of households carried out in each of the 624 selected EAs from September 2010 through January 2011. Maps were drawn for each of the clusters and all private households were listed. The listing excluded institutional living arrangements (e.g., army barracks, hospitals, police camps, and boarding schools).

All women age 15-49 and all men age 15-59 who were either permanent residents of the selected households or visitors who stayed in the household the night before the survey were eligible to be interviewed (CSA, 2011).

The structure of EDHS data and conceptual framework was shown as the following figure:

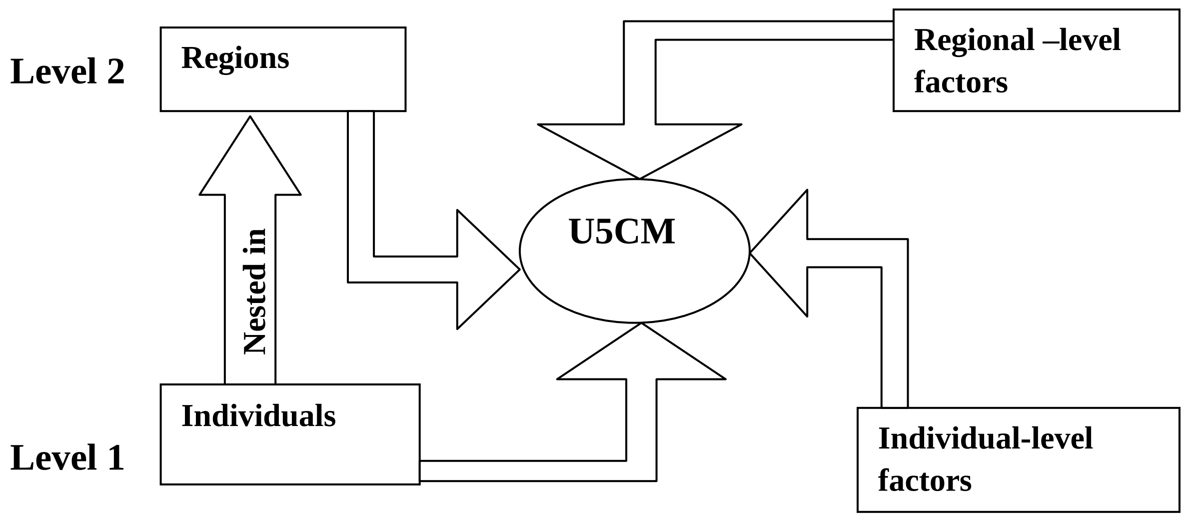

Fig 2.1: Structure of EDHS Data and its Conceptual Framework

2.3. Variables of the Study

The variables considered in this study taken based on earlier studies at the global and national level. As discussed in the literature review socio-economic, demographic and environmental characteristics are to be the essential and proximate determinants of child mortality at worldwide and national level as well. In this study, the potential determinant factors expected to be correlated with under-five child mortality are included as variables. Those variables considered in this study are classified as: dependent and explanatory or indicator variables stated below.

2.3.1. Dependent Variable

The dependent variable of interest for this study is child event before reaching five years of age, measured as the duration from birth to the age at death. Since in the DHS age at death (reported in days and months) is subject to heaping at certain ages, a discrete formulation of time is preferred to a continuous one. It is dichotomous coded as 1 if child died in the five years before the survey and 0 if alive.

2.3.2. Explanatory Variables

In the present study the following socio-economic, demographic and environmental factors which are expected to have impacts on under-five child mortality in Ethiopia are classified as individual level variables and regional level variables as given below:

Individual Level Variables

- Birth index
- Birth in last five years
- Birth order
- Breastfeeding status
- Child's sex
- Child size at birth
- Household wealth
- Mother's age at birth
- Mother's education
- Mother's work status
- Preceding birth interval
- Religion and
- Type of birth

Regional Level Variables

- Household toilet facility
- Place of birth delivery
- Place of residence
- Source of drinking water and
- Region

2.4. The Binary Logistic Regression Model

Logistic regression is a predictive model, like linear regression. But logistic regression involves prediction of a categorical dependent variable. The predictors can be continuous or dichotomous, as in regression analysis, but ordinary least squares regression (OLS) is not appropriate if the outcome is categorical. Whereas the OLS regression uses normal probability theory, logistic regression uses binomial probability theory. Binary logistic outcomes (dependent variables) are binary (dichotomous) and can be coded 0 (failure) and 1 (success). In the case of binary dependent variables, most assumptions in linear regression are violated and the dependent variable is restricted to the range of (0, 1), while in OLS regression there is no bounds. So, the solution for the violation of OLS assumption is logistic regression that does not make any assumption of normality, linearity, and homogeneity of variance for the independent variables. And by the log transformation will expand the range from (0, 1) to infinity.

Logistic regression is used to predict a categorical variable from a set of predictor variables are a mix of continuous and categorical variables or if they are not normally distributed (logistic regression makes no assumptions about the distributions of the predictor variables). While, discriminant function analysis is usually employed with categorical dependent variables if all of the predictors are continuous and normally distributed and logit analysis is usually employed if all of the predictors are categorical. Logistic regression has been especially popular with medical research in which the dependent variable is whether or not a patient dead. In general, it is appropriate to use binary logistic regression when the dependent variable is dichotomous (such as presence or absence, success or failure) (Hosmer and Lemeshow, 1989).

Binary logistic regression model is used to investigate the effect of predictors on the probability of having under-five child mortality is defined as follows:

- Dependent variable is given as:

$$Y_{ij} = \begin{cases} 1, & \textit{If child died before five years of age} \\ 0, & \textit{Otherwise} \end{cases} \tag{2.1}$$

$i = 1, 2, \ldots M \quad and \quad j = 1, 2, \ldots, N$

Where: M- is the number of under-five children in each region j.

N- is the number of region.

Let π denote the proportion of success (death of child before five years of age):

$$P(Y_{ij} = 1) = \pi_{ij}, \; P(Y_{ij} = 0) = 1 - \pi_{ij}$$

$$\text{And } Y_i \sim Bernoulli\ (\pi_i)$$

The logistic model is defined as follows. Let $X_{nx(k+1)}$ denote the single level binary logistic regression data matrix of k predicator variables of the under-five child death is given as:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1k} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & \cdot & x_{nk} \end{bmatrix} \sim nx(k+1), \; \beta = \begin{bmatrix} \beta_o \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \sim (k+1)x1 \tag{2.2}$$

X -is the design matrix

β - is the vector of unknown coefficients of the covariates and intercept

Then, the logistic regression function is given as:

$$\pi_i = \frac{e^{\beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}}}{1 + e^{\beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}}} = \frac{e^{X_i \beta}}{1 + e^{X_i \beta}} \tag{2.3}$$

Where: $\pi_i \ (i = 1, 2, \dots, n)$ is the probability of i^{th} child having death before five years of age given the vector of predictors (X_i).

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

$$\left[\frac{p(Y=1 \mid X_i)}{(1-p(Y=1 \mid X_i))}\right] = \left[\frac{\pi}{1-\pi}\right] = \exp(\beta_o + \beta_1 X_1 + \dots + \beta_k X_k)$$

Or

$$\log\left[\frac{p(Y=1 \mid X_i)}{(1-p(Y=1 \mid X_i))}\right] = \left[\frac{\pi}{1-\pi}\right] = \beta_o + \beta_1 X_1 + \dots + \beta_k X_k = \sum_{j=0}^{k} \beta_j \chi_{ij} \qquad (2.4)$$

Where: $i = 1, 2, \dots, n; j = 0, 1, 2, \dots, k$

The coefficient is interpreted as the change in the log-odds of having child death before five years of age per unit change of corresponding continuous covariate. In case of categorical predictor variable, it is interpreted as the log-odds of having child death before five years of age with a given category compared to the reference category (Dayton, 1992).

2.4.1. Assumption of Binary Logistic Regression

As indicated in the above sections, the advantage of the logistic regression is that it has flexible assumptions as compared with discriminant analysis. There are, however, other assumptions one should consider for the efficient use of logistic regression as detailed in (Hosmer and Lemeshow, 1989).

- Linear relationship exists through logit transformation of the dependent variable.
- The dependent variable is categorical to have to outcome.

- The dependent variable may assume a distribution from an exponential family (e.g. binomial, Poisson, multinomial, normal); binary logistic regression assume binomial distribution of the response.
- The groups for the predictors must be mutually exclusive and exhaustive.
- Larger samples are needed than for linear regression because maximum likelihood coefficients are large sample estimates. A minimum of 50 cases per predictor is recommended.
- There should not be severe collinearity among predictor variables.

2.4.2. Estimation of Coefficients in Logistic Regression Model

Based on assumption mentioned above, the logistic regression wants to use maximum likelihood to estimation the unknown coefficients of logistic regression model.

2.4.2.1. Maximum Likelihood Estimation for Logistic Regression

The maximum likelihood estimation method is appropriate for estimating the logistic model parameters due to this less restrictive nature of the underlying assumptions stated above. Hence, in this study the maximum likelihood estimation technique is used to estimate parameters of the model.

Consider the logistic regression model equation (2.3):

Then, the likelihood function with n observations is:

$$L(\beta, Y = 1 \mid X_i) = \prod_{i=1} \left(\frac{\exp\left(X_i^{'}\beta\right)}{1 + \exp\left(X_i^{'}\beta\right)} \right)^{\sum_{i=1}^{n} y_i} \left[1 - \left(\frac{\exp\left(X_i^{'}\beta\right)}{1 + \exp\left(X_i^{'}\beta\right)} \right) \right]^{n - \sum_{i=1}^{n} y_i} \tag{2.5}$$

And the log-likelihood function is:

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

The k + 1 score functions of β for the logistic regression model cannot be solved analytically. It is common to use a numerical algorithm, such as the Newton-Raphson algorithm, to obtain the MLEs. The information in this case will be a $(p+1)x(p+1)$matrix of the partial second derivative l with respect to the parameters, β. The inverted information matrix is the covariance matrix for $\hat{\beta}$ (Collet, 1991).

2.4.2.2. The Odds Ratio

The odds ratio is defined as the ratio of the probability of the occurrence of an event to non-occurrence of an event (Wang, 2011).

In binary logistic regression analysis, odds ratio is the exponent of the estimated coefficient $\exp(\hat{\beta})$. For each continuous covariate let say j, $\exp(\hat{\beta}_j)$ is the predicted change in odds having under-five child mortality for a unit increase in predictor j variable (Dayton, 1992). In case of categorical predictor variable, $\exp(\hat{\beta})$ is the predicted change in odds having under-five child mortality for a given category of the predictor variable with respect to the reference category.

2.4.3. The Assessment of Goodness Fit of Logistic Regression Model

2.4.3.1. The Likelihood Ratio Test

The likelihood ratio chi-square$\left(G^2\right)$ statistic is the test statistic commonly used for assessing the overall fit of the logistic regression model. The likelihood ratio test, also called log-likelihood test, it is based on - $2LL$ (-2 times log likelihood). The likelihood ratio statistic is obtained by subtracting the two times log likelihood ($-2LL$) for the final (full) model from the log likelihood for the intercept only model. This log likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the intercept only model L_0 over the maximized value of the likelihood function for the full model L_1. The likelihood test statistic is given by

$$G^2 = -2\log\left(\frac{L_0}{L_1}\right) = -2\left[\log\left(L_0\right) - \log\left(L_1\right)\right] = -2[LL_0 - \left(-LL_1\right)] \tag{2.6}$$

Where LL_0 the log likelihood value of the model which is have the intercept term only and LL_1 is the log likelihood value of the full model. The likelihood ratio statistic has a chi-square distribution and it tests the null hypothesis says all logistic regression coefficients except the constant are zero. The degrees of freedom are obtained by differencing the number of parameters in the both model. It compared with chi-square value at the difference between degree of freedom of both model. And p-value indicates that the probability of the deviance based on chi-square is greater than the tabulated chi-square. If p-value is less than 5 % level of significant leads the rejection of the null hypothesis that all the predictor effects are zero. When this likelihood test is significant, at least one of the predictors is significantly related to the response variable.

An alternative method for checking goodness of fit for individual binary data has been proposed by Hosmer and Lemeshow (2000) given below.

2.4.3.2. The Hosmer-Lemeshow Test Procedure

The test statistic for this test procedure is formulated under the null hypotheses that the model fits the data, and the alternative is that the model does not fit. The test statistic is constructed by grouping the data set into roughly 10 (g) groups (Hosmer and Lemeshow, 2000). The groups are formed by ordering the existing data by the level of their predicted probabilities. So the data are first ordered from least likely to have the event to most likely for the event. Then g (often 10) roughly equal size groups are formed. From each group the observed and expected numbers of events are computed for each group. The test statistic is

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - E_k)^2}{v_k} \tag{2.7}$$

Where, O_k and E_k are the observed and expected number of events in the k^{th} group, and v_k is a variance correction factor for the k^{th} group. If the observed number of events differs from what is expected by the model, the statistic $\hat{c}$ will be large and there will be evidence against the null hypothesis. This statistic has an approximate Chi-Squared distribution with (g – 2) degrees of freedom.

2.4.3.3. The Wald Test

The Wald statistic is an alternative test which is commonly used to test the significance of individual logistic regression coefficients for each independent variable. The hypothesis to be tested is: $H_0 : \beta_j = 0 \; vs \;\; H_A : \beta_j \neq 0 \quad j = 1, 2, ..., k$ at α level of significance.

The Wald test statistic, Z, for this hypothesis is

$$Z^2 = \left[\frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)} \right]^2 \approx \chi^2(1) \tag{2.8}$$

The Wald test is one of a number of ways of testing whether the parameters associated with a group of explanatory variables are zero. If the Wald test is significant for a particular explanatory variable then we would conclude that the parameters associated with these variables are not zero so that the variables should be included in the model otherwise the explanatory variables can be omitted from the model (Agresti, 1990).

2.4.3.4. R- Square Statistic

A number of measures have been proposed in logistic regression as an analog to R-square in multiple linear regressions. In logistic regression, there is no true R^2 value as there is in OLS regression. The maximum value that the Cox & Snell R-square attains is less than 1. The Nagelkerke R-square is an adjusted version of the Cox & Snell R-square and covers the full range from 0 to 1, and therefore it is often preferred (Bewick and Jonathan, 2005).

In SPSS, there are two modified versions of this basic idea, one developed by Cox & Snell and the other developed by Nagelkerke (Long, 1997) and (O'Connel, 2006). The Cox and Snell R-square is computed as follows:

Cox & Snell Pseudo-R^2 $$R^2 = 1 - \left[\frac{-2LL_{null}}{-2LL_{full}}\right]^{2/n}$$

Because this R-squared value cannot reach 1.0, Nagelkerke modified it. The correction increases the Cox and Snell version to make 1.0 a possible value for R-squared.

Nagelkerke Pseudo-R^2 $$R^2 = \frac{1 - \left[\frac{-2LL_{null}}{-2LL_{full}}\right]^{2/n}}{1 - \left(-2LL_{null}\right)^{2/n}}$$

Where: LL_{null} is log-likelihoods of the null model or the logistic model with just the constant

LL_{full} is log-likelihoods of the full logistic regression model or the logistic regression model contains all the k predictors.

2.4.4. Checking Multicollinearity, Outliners and Influential Cases

Multicollinearity Diagnostics

First, one has to check for multicollinearity before analyzing the data using binary logistic regression. Tolerance and VIF scores are not available through the logistic regression command, one way to compute these values is through the linear regression command, using one of the continuous variables assumed to be an indicator of under-five child mortality as the dependent variable and the rest of indicators including the response variables (under-five child mortality) as independent variables (Leech, 2005).

Similarly, in the correlation matrix for this case, it is not so easy to spot where the multicollinearity is? Another drawback with the correlation matrix is that multicollinearity between one variable with a combination of variables, will not be shown. A simple but sometimes subjective technique is to inspect the magnitude of the standard error (SE) of each variable. The SEs is very large implying multicollinearity exists and the model is not statistically stable. To “solve” this issue, start omitting the variable with high collinearity (Chan, 2004). There is no fixed criterion on how small the SE should be but it is a matter of judgment. However, correlation matrix has drawback it is better way to identify correlated variables in the study. To confirm multicollinearity diagnosis, it better to drop relatively correlated variables from the analysis.

2.5. Multilevel Logistic Regression Model

Before going to multilevel modeling, ones needed to go beyond the classical setup of a data Y and a matrix of predictors X stated in equation (2.2). The multilevel data structures with an observational study of the impacts of each indicators on under-five child mortality. The treatment is at the groups (region) level, but the outcome is measured on individual families.

The fact that the regional states in Ethiopia had a variety of environmental factors, health service provider, level of education of the people living in the community, level of educated family, access to safe drinking water, sanitation and different infrastructures to encourage the reduction of under-five child mortality at their region and national level. Indeed, not only regional-level differentials but also there are the individual-level factors attributed for under-five child mortality in addition to demographic factors of children as well. This differential among individual, region, national and also through continent level indicated the facts that, the rate of child mortality in developed and developing country has different structure. But, so many studies in single level (eliminate those variation across regional states) regarding under-five child mortality in the world wide and at national level that invites errors. In fact, there is clear heterogeneity among the individual and regional-level characteristics that leads to variations while clustered those factors at single level.

In the present study, multilevel binary logistic regression model was adopted to model under-five child mortality variations among regional states of Ethiopia. This study, started to built multilevel modelling of the variations for the impacts of individual and community (regional)-level on under-five child mortality starting from empty, random intercept and random coefficient binary logistic regression model as discussed as follows. First, ones better to check where there is heterogeneity proportion of under-five child mortality between regions in Ethiopia before going to multilevel analysis.

2.5.1. Heterogeneity Proportion

The basic data structure of the two-level regression is a collection of N groups ('units at two levels' or 'regions'), with in group $j, (j = 1, 2, \ldots, N)$ random sample of n_j level-one units ('individual' or 'number of under-five children living in the region j').

Consider the outcome variable in equation (2.1), Y_{ij} $(\mathrm{i} = 1, 2, \ldots, \mathrm{n_j};\ \mathrm{j} = 1, 2, \ldots, \mathrm{N})$ and denoted by for level-one unit i nested in level-two group j.

And the total sample size is $M = \sum_{j=1}^{N} n_j$. If one does not take explanatory variables into account, the probability of success is assumed constant in each group (Snijders and Bosker, 1999). Let the probability of having under-five child death in region j be denoted by π_j. The dichotomous outcome variable for the child i in region j, Y_{ij} can be expressed as the sum of the probability in region j, π_j (the average proportion of i levels in region j, $E(Y_{ij}) = \pi_j$) plus some individual dependent residual, that is

$$y_{ij} = \pi_j + \varepsilon_{ij} \tag{2.9}$$

The residual term is assumed to have mean zero and variance,

$$\operatorname{var}(\varepsilon_{ij}) = \pi_j (1 - \pi_j)$$

Since the outcome variable is coded 0 and 1, the group (region) sample average is the proportion of successes in group j given by:

$$\hat{\pi}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ij}$$

Where: $\hat{\pi}_j$ - is an estimate for the group-dependent probability π_j. Similarly, the overall sample average is the overall proportion of successes, π and is given by:

$$\pi = \frac{1}{M}\sum_{j=1}^{N}\sum_{i=1}^{n_j} Y_{ij}$$

2.5.2. Test of Heterogeneity Proportion

For the proper application of multilevel analysis the first logical step is to test heterogeneity of proportions between groups. Here we present two commonly used test statistics that are used to check for heterogeneity (Snijders and Bosker, 1999). To test whether there are indeed systematic differences between the groups, the well known Chi-Square test for contingency table can be used. In this case the Chi-Square test statistic is:

$$\chi^2 = \sum_{j=1}^{N} n_j \frac{(\hat{\pi}_j - \hat{\pi})^2}{\hat{\pi}(1-\hat{\pi})} \sim \chi^2(N-1) \tag{2.10}$$

It can be tested a chi-square distribution with $N-1$ degrees of freedom. This chi-squared distribution is an approximation valid if the expected number of success $(n_j\pi_j)$ (and of failures $(n_j(1-\pi_j))$ in each group all are at least one while 80 percent of them are at least 5 (Agresti, 1990).

2.5.3. Estimations of Between and Within Group Variance

The true variance between the group dependent probabilities (Snijders and Bosker, 1999), i.e. the population values of π_j, is given by:

$$\hat{\tau}^2 = S^2_{between} - \frac{S^2_{within}}{\tilde{n}} \tag{2.11}$$

Where: $\tilde{n}$ is defined as: $\tilde{n} = \frac{1}{N-1}\left\{M - \frac{\sum_{j=1}^{N} n_j^2}{M}\right\}$,

$$S_{between}^2 = \frac{\hat{\pi}(1-\hat{\pi})}{\tilde{n}(N-1)} X^2 \text{ and } S_{within}^2 = \frac{1}{M-N}\sum_{j=1}^{N} n_j(1-\pi_j),$$

Where: χ^2 is given in equation (3.10)

2.5.4. The Empty Model

The empty two-level model for a dichotomous outcome variable refers to a population of groups (level-two units) and specifies the probability distribution for group-dependent probabilities π_j (probability of having i^{th} child in j^{th} group (region) died before five year of age) then, consider equation (2.9) without taking further explanatory variables into account. We focus on the model that specifies the transformed probabilities $f(\pi_j)$ to have a normal distribution. This is expressed, for a general link function $f(\pi)$, by the formula

$$f(\pi_j) = \beta_0 + U_{oj}$$

Where $f(\pi_j)$ - is the population average of the transformed probabilities β_o and U_{oj} is the random deviation from this average for group *j*. If $f(\pi)$ is the logit function, then $f(\pi_j)$ 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:

$$Logit(\pi_j) = \beta_0 + U_{oj}$$

For the deviations U_{oj} it is assumed that they are independent random variables with a normal distribution with mean zero and variance δ_0^2. This model does not include a separate parameter for the individual level variance (Snijders and Bosker, 1999). This is because the individual level residual variance of the y_{ij} (death or alive of under-five children follows Bernoulli distribution directly from the probability of having under-five child death (π_j) which is given by:

$$\text{var}\left(\varepsilon_{ij}\right) = \pi_j\left(1 - \pi_j\right)$$

Denote by π_o the probability corresponding to the average value β_o , as defined by

$$f\left(\pi_o\right) = \beta_o$$

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

$$\pi_o = Logit\ \left(\beta_o\right) = \frac{\exp\ \left(\beta_o\right)}{1 + \exp\ \left(\beta_o\right)} \tag{2.12}$$

Because of the non-linear nature of the logit link function, there is no a simple relation between the variance of probabilities and the variance of the deviations U_{oj} (Snijders and Bosker, 1999). According to Snijders and Bosker (1999) there is an approximate formula, however, valid when the variances are small. The approximate relation (valid for small δ_o^2) between the population variance is:

$$\text{var}\left(\pi_j\right) \approx \frac{\delta_o^2}{\left(f'\left(\pi_o\right)\right)^2}$$

For the logit function, this yields:

$$\text{var}\left(\pi_j\right) = \left(\pi_o\left(1 - \pi_o\right)\right)^2 \delta_o^2$$

Note that an estimate of population variance $\text{var}\left(\pi_j\right)$ can be obtained by replacing sample estimates of π_o and δ_o^2.

2.5.5. Random Intercept Binary Logistic Regression Model

With grouped data, a regression that includes indicators for groups is called a varying-intercept model because it can be interpreted as a model with a different intercept within each group (Gelman and Hill, 2006). In this case the random intercept model is consider only random effect of each indicators of under-five child mortality meaning that the region differ with respect to the average value of under-five child death , but there is no different relation between indicators of under-five child mortality among groups (regional states).

A assume that X is predictor's data matrix denoted by: $X_h, (h = 1, 2, ..., k)$ these variables are denoted by with their values indicated by x_{hij} (Snijders and Bosker, 1999). Some or all of those variables could be level one variables, the success probability is not necessarily the same for all individual in a given group (region). From the above probability of having under-five child death depend on indicators was denoted by π_j. The outcome variable is split into an expected value and residual as in equation (2.9).

Then, random intercept model expresses the log-odds, i.e. the logit of π_{ij}, is the sum of a linear function of all indicators of under-five child mortality is given as:

$$Logit\ (\pi_{ij}) = \beta_{oj} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + ... + \beta_k x_{kij} = \beta_{oj} + \sum_{h=1}^{k} \beta_h x_{hij} \quad (2.13)$$

Where, logit(π_{ij}) does not include a level-one residual because it is an equation for the probability of having under-five child death (π_{ij}) rather than for the outcome y_{ij}

.

- β_{oj} is assumed to vary randomly and is given by the sum of an average intercept β_o and group (region) dependent deviations U_{oj} is given:

By replacing $\beta_{oj} = \beta_o + U_{oj}$ in equation (2.13)

We have:

$$Logit\ (\pi_{ij}) = \beta_o + \sum_{h=1}^{k} \beta_h x_{hij} + U_{oj}$$

Or
$$\pi_{ij} = \frac{\exp\left(\beta_o + \sum_{h=1}^{k} \beta_h \, x_{hij} + U_{oj}\right)}{1 + \exp\left(\beta_o + \sum_{h=1}^{k} \beta_h \, x_{hij} + U_{oj}\right)} \qquad (2.14)$$

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

U_{oj} - is random part of the model and It is assumed that they are mutually independent and normally distributed with mean zero and variance δ_o^2.

2.5.6. Random Slope Binary Logistic Regression Model

The multilevel modeling strategy accommodates the hierarchical nature of the DHS data and corrects the estimated standard errors to allow for clustering of observations within units (Goldstein, 2003). A significant random effect may represent factors influencing the outcome variable that cannot be quantified in a large-scale social survey. A random effects model thus provides a mechanism for estimating the degree of correlation in the outcome that exists at the region level, while also controlling a range of all indicators may potentially influence the outcome.

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

$$\beta_{oj} = \beta_o + U_{oj}$$

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

Thus, by substituting in equation (3.13) then, logit((π_{ij})) is given as:

$$Logit\ (\pi_{ij}) = (\beta_o + U_{oj}) + (\beta_1 + U_{1j})X_{1ij} = \beta_o + \beta_1 X_{1ij} + U_{oj} + U_{1j}X_{1ij} \quad (2.15)$$

Now, we have two random effects at group level, the random intercept U_{oj} and the random slope U_{1j}. It assumed that both random effects have mean zero. And the variances are denoted by δ_o^2, δ_1^2 and their covariance is δ_{o1}^2.

Where, β_o - is the average intercept of the response variable.

β_1 - is fixed regression coefficient given explanatory variable X_1.

U_0 - is the random coefficient in the model.

$U_o + U_1 X_{1ij}$ - is the random part of the model can be considered as interaction by group and predictors (X).

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

$$\operatorname{var}(U_{oj}) = \delta_{oo} = \delta_o^2$$
$$\operatorname{var}(U_{1j}) = \delta_{11} = \delta_1^2$$
$$\operatorname{cov}(U_{oj}, U_{1j}) = \delta_{o1}^2$$

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

That is:

$\beta_{oj} = \beta_o + U_{oj}, \beta_{1j} = \beta_1 + U_{1j}, ..., \beta_{hj} = \beta_h + U_{hj}, \quad for\ h = 1, 2, ..., k$, then we have:

$$Logit\ (\pi_{ij}) = (\beta_o + U_{oj}) + (\beta_1 + U_{1j})X_{1ij} + ... + (\beta_h + U_{hj})X_{hij}$$

$$= \beta_o + \sum_{h=1}^{k} \beta_h X_{hij} + U_o + \sum_{h=1}^{k} U_{hj} X_{hij} \tag{2.16}$$

Where, $\beta_o + \sum_{h=1}^{k} \beta_h X_{hij}$ - is fixed part of the model and

$U_o + \sum_{h=1}^{k} U_{hj} X_{hij}$ - is the random part of the model

$U_{oj}, U_{1j}, \ldots, U_{hj}$ - are assumed to be independent between groups but may be correlated within groups. So the components of the vector $U_{oj}, U_{1j}, \ldots, U_{hj}$ are independently distributed as a multivariate normal distribution with zero mean vector and variances and co-variances Ω given by:

$$\Omega = \begin{bmatrix} \delta_o^2 & \delta_{1o} & . & . & . & \delta_{ko} \\ \delta_{o1} & \delta_1^2 & . & . & . & \delta_{k1} \\ \delta_{o2} & \delta_{12} & \delta_2^2 & & & \delta_{k2} \\ . & & & . & & . \\ . & & & & . & . \\ \delta_{ok} & \delta_{1k} & . & . & . & \delta_k^2 \end{bmatrix}$$

2.5.7. Maximum Likelihood Estimation via Quadrature

The most common methods for estimating the parameter of multilevel logistic models are Marginal Quasi Likelihood (Goldstein, 1991; Goldstein and Rasbash, 1996), Penalized Quasi Likelihood (Laird, 1978; Breslow and Clayton, 1993). The numerical integrations approach and Laplace approximation seem to produce statistically more satisfactory estimates than MQL and PQL approaches.

The marginal likelihood is the joint probability of all observed responses given the observed covariates. For linear mixed models, this marginal likelihood can be evaluated and maximized relatively easily (Rabe-Hesketh and Skrondal, 2012). However, in generalized linear mixed models, the marginal likelihood does not have a closed form and must be evaluated by approximate methods. Now, we will construct this marginal likelihood step by step for a random intercept logistic regression model with covariates X_j. The responses are conditionally independent given the random intercept U_j and the covariates X_j. Therefore, the joint probability of all the responses $y_{ij}(1,\ldots,n_j)$ for cluster j given the random intercept and covariate is simply the product of the conditional probabilities of the individual responses:

$$P\left(y_{ij},\ldots,y_{n_j j} \mid X_j, U_j\right)=\prod_{i=1}^{n_j} P\left(y_{ij} \mid X_j, U_j\right)=\prod_{i=1}^{n_j} \frac{\exp\left(\beta_0+U_j+\beta' X_j\right)^{y_{ij}}}{1+\exp\left(\beta_0+U_j+\beta' X_j\right)}$$

as specified by the logistic regression model. To obtain the marginal joint probability of the responses, not conditioning on the random intercept U_j (but still on the covariate X_j), we integrate out the random intercept.

$$P\left(y_{ij},\ldots,y_{n_j j} \mid X_j\right)=\int P\left(y_{ij},\ldots,y_{n_j j} \mid X_j, U_j\right)\phi\left(U_j;0,\delta_o^2\right)du_j \qquad (2.17)$$

Where, $\phi\left(U_j;0,\delta_o^2\right)$ is the normal density of U_j with mean 0 and variance δ_0^2. Unfortunately, this integral does not have a closed-form expression. The marginal likelihood is just the joint probability of all responses for all clusters. Because the clusters are mutually independent, this is given by the product of the marginal joint probabilities of the responses for the individual clusters (Rabe-Hesketh and Skrondal, 2012).

$$L\left(\beta_o,\beta,\delta_o^2\right)=\prod_{j=1}^{N} P\left(y_{ij},\ldots,y_{n_j j} \mid X_j\right)$$

This marginal likelihood is viewed as a function of the parameters β_o, β, and δ_0^2 (with the observed responses treated as given). The parameters are estimated by finding the values of β_o, β, and δ_0^2 that yield the largest likelihood. The search for the maximum is iterative, beginning with some initial guesses or starting values for the parameters and updating these step by step until the maximum is reached, typically using a Newton–Raphson or expectation-maximization (EM) algorithm.

The integral over U_j in (2.14) can be approximated by a sum of R terms with e_r substituted for U_j and the normal density replaced by a weight w_r for the r^{th} term, $r = 1, 2, \ldots, R$:

$$\prod_{j=1}^{N} P\left(y_{ij}, \ldots, y_{n_j j} \mid X_j\right) \approx \sum_{r=1}^{R} P\left(y_{ij}, \ldots, y_{n_j j} \mid X_j, U_j = e_r\right) w_r \tag{2.18}$$

Where, e_r and w_r are called Gauss–Hermite quadrature locations and weights, respectively. This approximation can be viewed as replacing the continuous density of U_j with a discrete distribution with R possible values of U_j having probabilities $P(U_j = e_r)$ (Rabe-Hesketh and Skrondal, 2012).

And the likelihood function for random slope multilevel logistic regression model is described as follows. Let the response vector consist of the entire elements y_{ij}. Assuming that the conditional distribution of y_{ij} given the random effect (U_j) are independent of each other, the conditional density of y_{ij} is given by:

$$f_{y_{ij}|u_j}\left(y_{ij} \mid U_j\right) \sim Bernoulli \ \left(\pi_{ij}\right) \tag{2.19}$$

The expected value of the Bernoulli distribution equals π_{ij}, after applying the specified link function, modeled as a linear function of the covariates. The distributions of the random effects are multivariate normal $(u_1, u_2, ..., u_N \sim N(0, \Omega))$ which are independent draws from multivariate normal distribution.

The likelihood approach is used to estimate the fixed and the random parameters of the model by treating the actual random effect U as nuisance parameters, and work with the marginal likelihood function which is given by:

$$L(\beta, \Omega) = \int f(Y \mid U; \beta) f(U; \Omega) d_U \tag{2.20}$$

Where, $f(Y \mid U; \beta)$ is the distributional function for response conditional on the random effect. Here $f(U_j; \Omega)$ is the distribution function for the random effects. For two-level logistic Bernoulli response model, where random effects are assumed to be multivariate normal and independent across units, the marginal likelihood function is given by:

$$L(\beta; \Omega) = \prod_j \int \prod_i \left\{ (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}} \right\} f(U_j; \Omega) d_{U_j} \tag{2.21}$$

$$\pi_{ij} = \{1 + \exp(-X_{ij} \beta_j)\}^{-1}, \ \beta_j = \beta + U_j$$

Where $f(U_{ij}; \Omega)$ is typically assumed to be the multivariate normal density and can be written in the form $\int_{-\infty}^{\infty} p(u_j) f(u_j) d_{u_j}$

Gauss-Hermite quadrature approximates an integral such as the above as

$$\int_{-\infty}^{\infty} p(v) e^{-v^2} d_v \approx \sum_{q=1}^{Q} p(x_q) w_q \tag{2.22}$$

Where $\sum_{q=1}^{Q} p(x_q w_q)$ is a Gauss-Hermite polynomial evaluated at a series of quadrature points indexed by *q*. This function is then maximized using a suitable search procedure over the parameter space.

If we consider the model with a single random intercept at level two have:

$$P(U_j) = \prod_j \frac{\exp(X_{ij}\beta + U_j)}{[1 + \exp(X_{ij}\beta + U_j)]^2},\ f(u_j) = \delta_u^2 \phi$$

Where ϕ -is the standard normal density. The standard quadrature method selects points centered on zero, but U_j is not centered at zero and we may therefore need a very large number of quadrature points to cover the range. We have therefore essentially approximated the posterior density by a normal density with the same mean and standard deviation. A solution is to use adaptive quadrature (Goldstein, 2011). Quadrature methods have been applied successfully to poisson, binomial and multinomial and ordered category models and have been implemented in software packages (SAS and STATA (xtlogit and xtmelogit). Nevertheless, successful quadrature, even with the adaptive method, will often require a large number of quadrature points and even in simple cases convergence can be difficult to achieve (Lesaffre and Spiessens , 2001). This becomes especially important when there are several random coefficients since the quadrature points will now be in several dimensions so that the number of points increases geometrically with the number of random coefficients.

2.5.8. Multilevel Binary Logistic Regression Model Comparison

Deviance based on Chi-square

The deviance based on chi-square value for two models is obtained as two times the difference of log likelihood value of the two models. It is compared with the probability of deviance based on chi-square, is greater than critical value distributed to chi-squared at the difference between numbers of parameter in two models degree of freedom. If P-value is less than 5% level of significance, suggesting that multilevel empty model is significant.

The basic concept underlying this procedure is to compare the maximum likelihood under an assumed model with that of a baseline model. Let $\hat{L}_c$ be the maximized likelihood under the current model. This statistic cannot be used on its own to assess the lack of fit of the current model unless compared with a corresponding statistic of an alternative baseline model for the same data. This latter model is taken to be a model that fits the data perfectly. Such a model will have the same number of unknown parameters as there are observations. The model is termed the full or saturated model and the maximized likelihood under it is denoted by $\hat{L}_f$. The saturated model does not condense the information in the bulk of data into a simple summary, as it is not parsimonious. However, the maximum likelihood under this model is an intuitively appealing reference by which a corresponding value of a given model can be compared to assess the adequacy of the given model.

Let the statistic D, be defined as:

$$D = -2\log\left(\hat{L}_c \mid \hat{L}_f\right) = -2\left[\log\ \hat{L}_c - \log\ \hat{L}_f\right] \tag{2.23}$$

Large values of D are encountered when $\hat{L}_c$ is small relative to $\hat{L}_f$, indicating that the current model is a poor one. On the other hand, small values of D are obtained when $\hat{L}_c$ is similar to $\hat{L}_f$, indicating that the current model is a good one. The statistic D has chi-square distribution at degree of freedom equals to the difference between the number of parameter in full model and current model therefore, it measures the extent to which the current model deviates from the full model and is termed the deviance.

3. Results and Discussions

- In this study data from 10156 under-five children are included.
- About 6.9% of under-five children in Ethiopia are died before five years of age.
- Based on region of residence under-five child mortality rate 2.57% was the minimum in Addis Ababa.
- The maximum under-five child mortality rate was about 8.49% in Benishangul-Gumuz region.
- Under-five child mortality rates were 2.26% and 7.23%, for children having higher educational level and no educated mothers respectively.
- The rates of under-five child mortality were 8.11% and 5.91%, for children having mother age at birth 15-24 and 25-34 respectively.
- The rates of under-five child mortality were 4.46% and 35.7%, for children with birth index 1 and 4 respectively.
- The rates of under-five child mortality were 6.14% and 9.76% for children with very small and large size at birth respectively.

Some Variables	Number of Children	U5CM	Percentage of U5CM
Total U5CM	10156	704	6.93
Region			
Addis Ababa	388	10	2.57
Benishangul-Gumuz	895	76	8.49
Mother educational level			
No education	6955	503	7.23
Primary	2666	182	6.82
Secondary	358	15	4.19
Higher	177	4	2.26
Birth index			
Child with index 1	6911	308	4.46
Child with index 2	2791	305	10.9
Child with index 3	426	81	19
Child with index 4	28	10	35.7
Size of child at birth			
Very small	2233	137	6.14
Very large	1732	169	9.76
Mother age at birth			
15-24	2455	199	8.11
25-34	5280	312	5.91
35-44	2191	170	7.76
44+	230	23	10

Binary Logistic Regression Analysis

- The likelihood ratio test of overall model indicated that, there were significant relationships between variables which significant associated with under-five child mortality.
- The Nagelkerke R square is found 25.6% indicating that, those variables that had significant association with under-five child mortality included in binary logistic regression analysis are useful in predicting under-five child mortality and to indicate its variations among regional states.
- Hosmer and Lemeshow test is found to be statistically insignificant.

  i.e. Do not reject null hypothesis of the model fits the data very well.
- It indicates that binary logistic regression model of under-five child mortality fits the Ethiopian Demographic and Health data very well.

Table 4.2: Likelihood Ratio Test of Overall Model

Model	-2*Log likelihood	df	X^2	P-value
Empty	5116.04	1	1086.4	0.000
Full	4029.631	29		

df – is degree of freedom

Table 4.4: Model Summary of Binary Logistic Regression Model

Step	-2*Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4029.631	0.101	0.256

Table 4.5: Hosmer and Lemeshow Test of Goodness Fit

Step	Chi-Square	df	P-value
1	12.134	8	0.145

Results of Binary Logistic Regression Analysis of Under-five Child Mortality

- Regions, Mother educational level, birth index, child size at birth, mother age at birth, type of birth and breastfeeding status have significant impacts on under-five child mortality.
- The odds of child living in all regional states except in Dire Dawa, being died before five years of age are higher than that of living in Addis Ababa.
- The odds of child having higher educated mothers being died before five years of age, is reduced by 68.4% as compared to child having none educated mothers. The odds of children with birth index 2, 3 and 4 being died before five years of age are higher as compared to children with birth index 1.
- The odds of being died before five years of age are reduced by 35.3% for children having mothers with age at birth between 25 and 34 as compared to that having mothers age at birth between 15 and 24.
- The odds of being died before five years of age for children with multiple births are higher than that of single birth.
- The odds of being died before five years of age for breastfed child are reduced by 96.7% as compared to that of never breastfed.

Variables	β	S.E.	Wald	df	P-value	Exp(β)	95% CI. for Exp(β)	
							Lower	Upper
Region **(ref: Addis Ababa)**			30.57	9	0.000*			
Tigray	1.122	0.40	7.67	1	0.006*	3.071	1.388	6.795
Affar	1.451	0.40	12.95	1	0.000*	4.267	1.936	9.402
Amhara	1.2	0.40	8.762	1	0.003*	3.322	1.5	7.355
Oromiya	0.934	0.4	5.458	1	0.019*	2.545	1.162	5.571
Benishangul-Gumuz	1.221	0.41	8.881	1	0.003*	3.39	1.519	7.568
SNNPR	1.304	0.4	10.64	1	0.001*	3.686	1.683	8.071
Gambela	0.819	0.41	3.891	1	0.049*	2.269	1.005	5.12
Harari	0.921	0.42	4.87	1	0.027*	2.512	1.109	5.692
Dire Dawa	0.642	0.42	2.306	1	0.129	1.901	0.83	4.353
Mother Education **(ref: No Education)**			5.874	3	0.118			
Primary	-0.05	0.11	0.198	1	0.657	0.951	0.764	1.185
Secondary	-0.49	0.33	2.312	1	0.128	0.609	0.321	1.154
Higher	-1.15	0.57	4.107	1	0.043*	0.316	0.103	0.963
Wealth index **(ref: Lowest)**			2.71	4	0.608			
Second	0.014	0.13	0.011	1	0.917	1.014	0.784	1.31
Middle	-0.04	0.14	0.077	1	0.781	0.963	0.736	1.259
Fourth	-0.21	0.15	2.002	1	0.157	0.813	0.61	1.083
Highest	0.02	0.21	0.009	1	0.925	1.02	0.676	1.54
Birth index **(ref: Child with index 1)**			112.5	3	0.000*			
Child 2	0.714	0.09	55.49	1	0.000*	2.043	1.693	2.465
Child 3	1.226	0.16	56.05	1	0.000*	3.408	2.472	4.698
Child 4	2.68	0.44	37.31	1	0.000*	14.58	6.171	34.447
								Cont…

Child Size at Birth **(ref: Vey Small)**			38.57	4	0.000*			
Very larger	0.502	0.14	12.69	1	0.000*	1.653	1.254	2.179
Larger than average	0.579	0.15	14.94	1	0.000*	1.785	1.331	2.394
Average	-0.02	0.13	0.017	1	0.896	0.984	0.768	1.259
Smaller than average	-0.17	0.19	0.807	1	0.369	0.84	0.574	1.229
Mother Age at birth **(ref: 15-24)**			22.03	3	0.000*			
25-34	-0.43	0.11	15.62	1	0.000*	0.647	0.521	0.803
35-44	-0.11	0.13	0.79	1	0.374	0.893	0.696	1.146
44+	0.239	0.26	0.845	1	0.358	1.269	0.763	2.111
Type of birth **(ref: Single)**								
Multiple	1.473	0.17	78.11	1	0.000*	4.363	3.147	6.048
Breastfeeding status **(ref: Never breastfed)**								
Breastfed	-3.41	0.13	670.0	1	0.000*	0.033	0.025	0.043
Place of residence **(ref: Rural)**								
Urban	0.114	0.21	0.291	1	0.59	1.12	0.741	1.694
Constant	-0.93	0.42	4.83	1	0.028*	0.394		

(* Significant at 5% level) and (ref - is reference category)

Multilevel Binary Logistic Regression Analysis

Chi-square Test of heterogeneity

- It indicates that there is heterogeneity of under-five child mortality between regional states.

Result of Empty Model

- It indicated that under-five child mortality variations among regional states of Ethiopia was non-zero.
- And about 1.156% of the variance in under-five child mortality at individual level could be attributed to differences across regional states.

U5CM	β	Std. Err.	Z	P>z	[95% CI. Interval]	
Fixed effect					**Lower**	**Upper**
β_0 = Intercept	-2.641	0.071	-33.8	0.000*	-2.79	-2.488
Random part						
Sigma (δ^2_{uo}) =Variance	0.1962	0.086	2.27	0.023*	0.083	0.465
Intra-region correlation coefficient						
ICC (Rho (ρ_u))	0.0116	0.01	4.81	0.028*	0.002	0.062
Likelihood-ratio test of rho=0: chibar2(01) = 4.81 Prob >= chibar2 = 0.014*						

(* Significant at 5%) and (ICC - intra-region correlation coefficient)

Random Intercept Binary Logistic Regression Analysis

- The variance of the random component related to intercept term is found to be significant.
- Indicating that, under-five child mortality variations among regional states of Ethiopia was non-zero.
- Mothers educational level, birth index, child size at birth, mother age at birth, type of birth and breastfeeding status have significant impacts and contribution to under-five child mortality variations among regional states of Ethiopia.

Variables	B	Std. Err.	Z	P>z	Odds	[95% CI]	
Fixed part						**Lower**	**Upper**
Mother Education Level (ref: No education)							
Primary	-0.063	0.111	-0.58	0.565	0.938	0.754	1.166
Secondary	-0.546	0.323	-1.69	0.091	0.578	0.307	1.09
Higher	-1.31	0.571	-2.29	0.022*	0.269	0.088	0.826
Wealth index (ref: Lowest)							
Second lowest	0.001	0.129	0.01	0.992	1	0.776	1.29
Middle	-0.058	0.135	-0.43	0.665	0.943	0.723	1.229
Fourth	-0.219	0.144	-1.52	0.129	0.8	0.604	1.066
Higher	-0.042	0.209	-0.2	0.84	0.96	0.635	1.446
							Cont...

Birth index (ref: Child with index 1)							
Child 2	0.716	0.096	7.48	0.000*	2.046	1.696	2.468
Child 3	1.222	0.163	7.49	0.000*	3.396	2.466	4.677
Child 4	2.669	0.437	6.1	0.000*	14.435	6.122	34.037
Child size at birth (ref: Very Smaller)							
Very larger	0.492	0.14	3.51	0.000*	1.636	1.24	2.154
Larger than average	0.565	0.149	3.79	0.000*	1.759	1.313	2.357
Average	-0.015	0.125	-0.12	0.906	0.985	0.77	1.26
Smaller than average	-0.189	0.194	-0.97	0.331	0.827	0.565	1.212
Month age at birth (ref: 15 up to 24)							
25-34	-0.44	0.109	-4.01	0.000*	0.643	0.518	0.798
35-44	-0.113	0.126	-0.9	0.369	0.892	0.696	1.144
44+	0.24	0.258	0.93	0.353	1.271	0.765	2.112
Type of Birth (ref: Single)							
Multiple	1.463	0.166	8.82	0.000*	4.319	3.12	5.979
Breastfeeding status (ref: Never breastfed)							
Breastfed	-3.384	0.13	-25.8	0.000*	0.034	0.026	0.044
Type of place of residence (ref: Rural)							
Urban	0.021	0.213	0.1	0.92	1.022	0.672	1.551
β_0 = Intercept	0.619	0.206	3	0.003*			
Random part							
Sigma (δ^2_{ou})	0.218	0.088	2.477	0.013*		0.098	0.485
ICC (Rho (ρ_u))	0.014	0.011	7.18	0.008*		0.003	0.066

(* Significant at 5%)
(ref - is reference category)
(ICC - intra-region correlation coefficient)

Multilevel Model Comparison

- Based on deviance based on Chi-square random intercept binary logistic regression model was the best fit model as compared to empty and random slope multilevel binary logistic regression models.
- The variations of under-five child mortality among regional states of Ethiopia were accounted only in the intercept term.
- However, those significant variables did not show any underline variations of under-five child mortality among regional states.
- The under-five child mortality variations among regional states were non-zero and its accounted by random intercept term only.

Model comparison statistics	**Empty model**	**Random intercept**	**Random coefficient model**
-2*log likelihood	5111.23	4054.82	4053.75
Deviance based on Chi –square	4.8138	1056.41	1.074
Degree of freedom (df)	2	22	27
P-value	0.0282*	0.000*	0.956
AIC	5115.23	4098.82	4107.75
BIC	5129.68	4257.79	4302.84

(* Significant at 5% level)

4. Conclusions

- Under-five child mortality was significantly associated with geographical region.
- The probability of under-five children living in all regional states except Dire Dawa, were more likely to die before five years of age than that of living in Addis Ababa.
- Under-five child mortality variations among regional states were accounted by the random intercept terms of the model.
- Mother educational level, birth index, child size at birth, mother age at birth, type of birth and breastfeeding status had significant impact and contribution to under-five child mortality variations among regional states.
- The probability of an under-five child having mother with higher educational level, being died before five years of age was less than that having mother with no education.
- The probability of a child with higher birth index, being dying before five years of age was higher than that with birth index 1.
- The probability of under-five child with larger than average size at birth, was more likely to die before five years of age than that with very small size at birth.
- The probability of under-five child born from mother with age at birth from 25 up to 34 was less likely to die before five years of age than that born from mother with age at birth from 15 up to 24.
- The probability of child with multiple births was more likely to die before five years of age than under-five child with single birth.

5. Recommendations

- Supporting mother's to educate themselves.
- Preferable if households have less birth index or less birth within five years.
- Improvement in maternal health care service will be appropriate to control larger size of child at birth.

- Let mothers preferable to give birth at ages between 25 up to 34.
- Multiple born children need professional cares and special attention of their parents.
- Mothers have to develop the culture of breastfeeding of children.
- Further studies should be conducted to identify others factors that affect and contribute to under-five child mortality variations among regions.
- Multilevel models are appropriate method that investigates the effects of demographic, socio-economic and environmental factors on under-five child mortality and to take into account its variations among regional states.

6. References

Agresti, A. (1990). Categorical Data Analysis. John Wiley and Sons, New York.

Antai, D., Ghilagaber, G., Wedrén, S., Macassa, G. and Moradi, T. (2009). Inequities in Under-Five mortality in Nigeria: Differentials by Religious Affiliation of the Mother. *J Relig Health*, Vol 48:P 290-304.

Bewick, L. and Jonathan, B. (2005). Statistics Review 14: Logistic Regression.

Chan, Y. H. (2004): Logistic Regression Analysis. Biostatistics 202, *Singapore Med J*, Vol 45(4): P149.

Collet, D. (1991). Modeling Binary Data, chapman and Hall, London.

CSA (2011 & 2012). Ethiopia Demographic and Health Survey data. Central Statistical Authority (Ethiopia) and ORC Macro; Addis Ababa, Ethiopia and Calverton, Maryland, USA.

Dayton, C.M. (1992). Logistic Regression Analysis. Department of Measurement, Statistics and Evaluation, University of Maryland.

Garenne, M. and Gakusi, E. (2006). "Health Transitions in sub-Saharan Africa: overview of mortality trends in children under 5 years old (1950-2000)." *Bull World Health Organ*, Vol 84(6):P 470-478.

Gelman, A. and Hill, J. (2006). Data Analysis Using Regression and Multilevel /Hierarchical Models. Columbia University, *Journal of the American Statistical Association.*

Goldstien, H. (2011): Multilevel Statistical Models (4th edition). University of Bristol, UK. Wiley series in probability and statistics, *John Wiley & Sons, Ltd.*

Goldstein, H. and Rasbash, J. (1996). Improved Approximations for Multilevel Models with Binary Responses. *J. Roy. Statist. Soc. A*, Vol 159:P 505–513.

Goro, I.M. (2007). The Stalling Child Mortality in Ghana: the case of the three northern regions. University of Cape Town, Cape Town, South Africa.

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

Kandala, N. B., Ji, C., Stallard, N., Stranges, S. and Cappuccio, F. P. (2007). Spatial Analysis of Risk Factors for Childhood Morbidity in Nigeria. *Am J Trop Med Hyg*, Vol 77:P 770-779.

Kumar, P.P. and Gemechis File (2010). Infant and Child Mortality in Ethiopia: a statistical analysis approach, Mathematics Dept. Jimma Univerity, Indian Expatriate.

Laird, N. M. (1978). *Empirical Bayes Methods for two-way Contingency tables. Biometrika*, Vol 65:P 581–590.

Leech, N.L., Barrett, K.C. and Morgan, G.A. (2005). SPSS for Intermediate Statistics: Use and Interpretation, Second Edition. University of Colorado at Denver, Colorado State University,*Lawrence Erlbaum Associates, Publishers;* Mahwah, New Jersey London.

Lindstrom, D.P. and Gebre-Egziabher, K. (2001). The Impact of Infant and Child Death on Subsequent Fertility in Ethiopia, Institute of Public Health Florida A&M University Tallahassee; FL 32307 850-412-7422.

Long, J.S. (1997). Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage.

Mahfouz, M.S., Surur, A.A, Ajak, D.A. and Eldawi, E.A.(2009). Level and Determinants of

Infant and Child Mortality in Malakal Town, Southern Sudan. *Sudanese journal of public health*, Vol 4(2).

MoFED (2010). Trends and Prospects for Meeting MDGs by 2015, MOH; Ethiopian MDGs Report, Addis Ababa, Ethiopia.

Mondal, N.I., Hossain, K. and Ali, K. (2009). Factors Influencing Infant and Child Mortality: A Case Study of Rajshahi District, Bangladesh, *J Hum Ecol*, Vol 26(1):P 31-39.

Omar, B.A, Lopez, A.D. and Inoue, M. (2000). The Decline in Child Mortality: a reappraisal. Bulletin of the World Health Organization, Vol 78:P 1175–1191.

Pickett, K.E. and Pearl, M. (2001). "Multilevel Analyses of Neighbourhood Socioeconomic Context and Health Outcomes: a critical review." *J Epidemiol Community Health*, Vol 55(2):P 111-122.

Rabe-Hesketh, S. and Skrondal, A. (2012). Multilevel and Longitudinal Modeling Using Stata (Third Edition). College Station, TX: Stata Press. Volume II: Categorical Responses, Counts, and Survival.

Siddiqi, A, Irwin, LG and Hertzman, C. (2007). The Region as a Sphere of Influence. Total environment assessment model for early child development: Evidence report World Health Organization's Commission on the Social Determinants of Health.

Snijders, T. A. B. and Bosker, R.J. (1999). An Introduction to Basic and Advanced Multilevel Modeling (1st edition). Department of Statistics, University of Poone.

Stephenson, R., Baschieri, A., Clements, S., Hennink, M., and Madise, N. (2006). "Contextual Influences on the use of Health Facilities for Childbirth in Africa." *American Journal of Public Health*, Vol 96(1):P 84-93.

United Nation (2011). The Millennium Development Report. Goals 4: Reduction of Child mortality, Addendum.

Wang, F. (2011). Logistic Regression: Use and Interpretation of Odds Ratio (OR). Surveillance and Assessment Branch, AHW, Community Health Sciences, the University of Calgary.

Zeleke Worku (2009). Factors That Affect Under-Five Mortality among South African Children. Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA.