Biomedical Statistics and Informatics
2017; 2(1): 27-36
http://www.sciencepublishinggroup.com/j/bsi
doi:10.11648/j.bsi.20170201.16

Multilevel Logistic Modelling of Adult Mortality Among Regional States in Ethiopia

Kassaye Wudu Seid[1, *], Ayele Taye Goshu[2]

[1]Department of Statistics, Arba Minch University, Arba Minch, Ethiopia
[2]Schools of Mathematical and Statistical Sciences, Hawassa University, Hawassa, Ethiopia

Email address:
kassayewudu16@gmail.com (K. W. Seid)
[*]Corresponding author

To cite this article:
Kassaye Wudu Seid, Ayele Taye Goshu. Multilevel Logistic Modelling of Adult Mortality Among Regional States in Ethiopia. *Biomedical Statistics and Informatics.* Vol. 2, No. 1, 2017, pp. 27-36. doi: 10.11648/j.bsi.20170201.16

Received: December 16, 2016; **Accepted**: December 27, 2016; **Published**: February 13, 2017

Abstract: The purpose of this study is to investigate adult mortality variations among regional states in Ethiopia. Data from Ethiopian Demographic and Health Survey (EDHS 2011) are used. Multilevel logistic regression model is fitted to the data. The findings show that the correlates of adult mortality are adult's age, age of household head, sex of household head, wealth index, type of toilet facility, and type of cooking fuels. The between-region variance is estimated to be 0.1727 which is significant at 5% level of significance, indicating that variation of adult mortality among regional states is non-zero. The intra-correlation is estimated to be 0.0099, suggesting that about 1% of the variation in adult mortality could be attributed to differences across regional states. The variance of the random coefficient model is statistically significant, implying the presence of adult mortality variation among regional states of the country. The deviance statistic is about 4113.945 and it is compared with Chi-square at difference of full and empty model with degree of freedom 33.924 at df 22 and p-value < 0.000. This reveals that there is enough evidence against null hypothesis. Therefore, the multilevel model consisting the explanatory variables is considered the final model. It revealed that the random coefficient multilevel logistic regression analysis suggests that there exist considerable differences in adult mortality across regions in the country.

Keywords: Adult Mortality, EDHS, Ethiopia, Heterogeneity, Logistic Regression Model, Multilevel Model, Random Effects, Regional Variation

1. Introduction

Many efforts have been made to study infant and child mortality along with maternal mortality in almost all developing countries including Ethiopia. But such academic attention has been lacking in studying the determinants of adult mortality in those countries. Very few studies are available on the consequences of adult mortality, particularly with regard to its socio-economic and health impacts on the family. Adults constitute the working force and very crucial for the development of any society because they are economically productive, biologically reproductive, and responsible for the support of children and elderly population. Excessive mortality of adults can influence various aspects of house hold behavior, since the loss of adults decreases the house hold income and thus human capital investments of next generation.

According to [1], global adult mortality rate has reduced from 0.198 in 1990 to 0.149 in 2015. Mortality rate for males is higher than that of females. Disease burden from non-communicable diseases among adults is increasing in developing countries due to aging and health transitions. So the level of adult mortality is an important indicator of the comprehensive assessment of the mortality pattern a population.

Demographic research on adult mortality is significant for understanding the health consequences of social inequality, human behavior, biological factors, and various other forces in human populations. In turn, mortality patterns can have profound influences on the size and composition of these populations. Thus, understanding adult mortality patterns is crucial to comprehending the dynamics of human society [2].

The study by [3] in Malawi reported that adult death due to external causes differs by sex, while there was no pattern with respect to age. They also reported drowning, uncontrolled epilepsy and alcohol as important risk factors.

This study deals with individual and regional determinant factors of adult mortality in Ethiopia during their productive age. Therefore, in this research we attempted to indicate some major demographic and socio economic determining factors of adult mortality in the country.

2. Materials and Methods

2.1. Data and Population

This study uses the 2011 Ethiopia Demographic and Health Survey (EDHS 2011). The EDHS 2011 was conducted in collaboration among the Ministry of Health, the Central Statistical Agency, and partner organizations. Data were collected from September 2010 through June 2011 with a nationally representative sample of nearly 18,500 households (CSA, 2011). All adults with age 15-49 in these households were considered. Therefore, this study excludes those individuals whose age were below 15 and 50 & above.

2.2. Variables Considered in the Study

The response variable in this study is adult mortality status in the country. Therefore, the response variable for the i[th] individual residing in j[th] region is represented by a random variable Y_{ij} with two possible values as 1 and 0. That is,

$$Y_{ij} = \begin{cases} 1, \text{if a person i residing in region j died during his/her productive age (15-49 years old)} \\ 0, \text{if alive till the survey time} \end{cases}$$

i= 1, 2, 3... n; j= 1, 2, 3..., k. where k is number of regions and total n number of individuals.

The explanatory variables in this study are:-

Individual-level factors:
- Age of adult
- Sex of adult
- Age of household head
- Household wealth index
- Sex of household head

Regional-level factors:
- Household toilet facility
- Place of residence
- Region
- Source of drinking water
- Types of cooking fuel

2.3. Sampling Design and Techniques

The EDHS 2011 sample was selected using a stratified, two-stage cluster design and EAs were the sampling units for the first stage. The sample included 624 EAs, 187 in urban areas and 437 in rural areas. Households comprised the second stage of sampling. A complete listing of households was carried out in each of the 624 selected EAs from September 2010 through January 2011. The listing excluded institutional living arrangements and collective quarters such as army barracks, hospitals, police camps, and boarding schools.

2.4. Statistical Models

2.4.1. Model Specification

The multilevel regression model is a standard statistical model enabling the study of simultaneous effects of factors, from individual adults as level one and regions as level two, on adult mortality. The analysis of a two level model involves different steps for estimation: null model, and random coefficients model. The null model involves no predictors while the final model involves random intercept and random slopes with predictor variables from the two levels.

The response variable Y_{ij} is dichotomous: 1 if an adult individual i residing in region j dies, and 0 otherwise. π_{ij} is the probability that the individual dies within the assumption of Bernoulli distribution. There are k explanatory variables: $X_{1ij}, X_{2ij}, \ldots, X_{kij}$ observed on the subject.

For the logistic regression model, the link function is logit(π_{ij}) or just log of odds $\pi_{ij}/(1-\pi_{ij})$.

2.4.2. The Null Model

The null mode is defined as:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + U_{0j} \quad (1)$$

The index i indicates individual (adult) for level one, j indicates region for level two, U_{0j} is level two errors, β_0 is the overall average of adult mortality rate.

Intra Class Correlation (ICC)

Within the assumption of Bernoulli distribution for the response, the variance becomes:

$$Var(Y_{ij}) = \pi_{ij}(1-\pi_{ij}) \quad (2)$$

It is, therefore, possible to decompose the variance at two levels - to assess how much of the variation is due to regions σ^2_u and how much is due to individuals differences σ^2_ε .

The intra class correlation (ICC) is given as follows:

$$ICC = \rho = \frac{\sigma^2{}_u}{\sigma^2{}_u + \sigma^2{}_\varepsilon} \quad (3)$$

This measure the correlation between variables on the same levels or proportion of group level variances compared to total variance.

2.4.3. The Multilevel Logistic Model

The classical logistic regression model involves fixed intercept, fixed slopes with explanatory variables. It can be expressed by:

$$\log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k \quad (4)$$

The logistic model assumes a linear relationship between the logit of the probability of death event and the linear function of the explanatory variables.

The multilevel logistic regression model involves random intercept, random slopes with explanatory variables and can be expressed by:

$$\log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + U_0 + (\beta_1 + U_1)X_1 + ... + (\beta_k + U_k)X_k \quad (5)$$

where each random effect U is assumed to have normal distribution with mean zero and contact variance. Both the classical and multilevel logistic models are fitted to the data.

2.4.4. Assessment of Goodness of Fit of Logistic Model

The Likelihood Ratio Test

The likelihood ratio chi-square (G^2) 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 (-2 times log likelihood). The overall significance is tested using what SPSS calls the Model Chi square, which is derived from the likelihood of observing the actual data under the assumption that the model that has been fitted is accurate. There are two hypotheses to test in relation to the overall fit of the model:

H_0: The model is a good fitting model.

H_1: The model is not a good fitting model (i.e. the predictors have no significant effects).

The difference between 2LL for the best-fitting model and 2LL for the null hypothesis model (in which all the β values are set to zero) is distributed like chi squared, with degrees of freedom equal to the number of predictors; this difference is the Model chi square. i.e

$$G^2 = -2\log(\frac{L_0}{L_1}) = -2(\log L_0 - \log L_1) = -2(LL_0 + LL_1) \quad (6)$$

where LL_0 is the log likelihood value of the model which has the intercept term only and LL_1 is the log likelihood value of the full model. It compared with chi-square value at the difference between degree of freedom of both model. If p-value is less than 5% level of significant it leads to 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.

The Hosmer-Lemeshow Test Procedure

An alternative to model chi-square is the Hosmer and Lemeshow test which divides subjects into 10 ordered groups of subjects [5]. And then compares the number actually in the each group (observed) to the number predicted by the logistic regression model (predicted). The 10 ordered groups are created based on their estimated probability; those with estimated probability below.1 form one group, and so on, up to those with probability.9 to 1.0. Each of these categories is further divided into two groups based on the actual observed outcome variable (success, failure). The expected frequencies for each of the cells are obtained from the model. A probability (p) value is computed from the chi-square distribution with (g-2) degrees of freedom to test the fit of the logistic model. If the H-L goodness of fit test statistic is greater than.05, as we want for well fitting models, we fail to reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the models estimates fit the data at an acceptable level. That is, well fitting models show non-significance on the H-L goodness of fit test. This desirable outcome of non significance indicates that the model prediction does not significantly differ from the observed.

$$C^2 = \sum(\frac{Ok - Ek}{Vk}) \quad (7)$$

where, Ok and Ek are the observed and expected number of events in the k^{th} group, and Vk 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 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.

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_1: $\beta_j \neq 0$, j=1, 2, 3... K at α level of significance. The Wald test statistic, Z, for this hypothesis is given as follows:

$$Z^2 = (\frac{\beta_j}{se(\beta_j)})^2 \to \chi^2{}_1 \quad (8)$$

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 [6].

R^2 Statistics

As we have seen above, having defined residuals for logistic regression, we can form the usual R^2 statistic, although it is rarely used. It is almost always rather low, since observed values need to be either 0 or 1, but predicted values are always in between these extremes. The maximum value that the Cox and Snell R-square attains is less than 1. The Nagelkerke R-square is an adjusted version of the Cox and Snell R-square and covers the full range from 0 to 1, and therefore it is often preferred [7]. In SPSS, there are two modified versions of this basic idea, one developed by Cox and Snell the other developed by Nagelkerke [8] - [9]. The Cox and Snell R-square is computed as follows:

$$R^2 = 1-(-\frac{2LL_0}{2LL_1})^{2/n} \tag{9}$$

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. Where: LLnull is log-likelihoods of the null model or the logistic model without any explanatory variable. LLfull is log-likelihoods of the full logistic regression model or the logistic regression model contains all the k predictors (explanatory variables).

2.4.5. Outliers, Influential Diagnostics and Multicollinearity

1. DFBETA(S) is a diagnostic measure which measures the change in the logit Coefficients for a given variable when a case is dropped. If DFBETAs is less than unity it implies no specific impact of an observation on the coefficient of a particular predictor variable, while DFBETA of a case is greater than 1.0, is considered as potential outliers.
2. Analog of Cook's influence statistics of a case greater than 1.0 indicates that a potential outlier, while the value of the leverage statistic less than one shows that no subject has a substantial large impact on the predicted values of a model.
3. Multicollinearity or simply co-linearity is a situation where there is close to a near perfect linear relationship among some or all of the independent variables in a regression model.
4. Variance Inflation Factor (VIF) and Tolerance are two measures that can guide a researcher in identifying multicollinearity. Before developing the concepts, it should be noted that the variance of the OLS estimator for a typical regression coefficient (say β_i) can be shown to be the following

$$Var(\beta_i) = \frac{\sigma^2}{S_{ii}(1-R^2)} \tag{10}$$

where $S_{ii} = \Sigma (X_{ij}$ - mean of $X_i)^2$ and R^2 is the unadjusted R^2 when we regress X_i against all the other explanatory variables in the model, that is, against a constant, X_2; X_3; X_{i-1}; $X_{i+1}\ldots;X_k$.

Suppose there is no linear relation between X_i and the other explanatory variables in the model. Then, R^2 will be zero and the variance of β_i will be δ^2/S_{ii}. Dividing this into the above expression for Variance of (β_i), we obtain the variance inflation factor and tolerance as

$$\begin{aligned} VIF(\beta i) &= \frac{1}{1-R^2} \\ Tolerance &= \frac{1}{VIF} = 1-R^2 \end{aligned} \tag{11}$$

It is readily seen that the higher VIF or the lower the tolerance index, the higher the variance of β_i and the greater the chance of finding β_i insignificant, which means that severe Multicollinearity effects are present. The procedure is to choose each right hand side variable (that is, explanatory variable) as the dependent variable and regress it against a constant and the remaining explanatory variables. We would thus get k-1 values for VIF. If any of them is high, then Multicollinearity is indicated. We say multicollinearity exists and sever if VIF >10 and moderate if 5<VIF<10, and no sever Co linearity if VIF<5.

2.4.6. Model Comparisons

Akaikes Information Criterion (AIC) is the expected estimated relative Kullback-Leibler (K-L) distance, where the K-L distance is the minimum distance between a model and full reality [10]. AIC it is given as:

$$AIC = -2\ln(L_{model}) + 2k \tag{12}$$

Where: K is number of estimated parameters and L (model) is the likelihood of the model.

Bayesian Information Criterion (BIC) is also known as the schwarz criterion after gideon Schwarz and virtually identical to the minimum description length criterion [10]. The formula is given as:

$$BIC = -2\ln(L\,model) + k * \log(n) \tag{13}$$

where:

- k is number of estimated parameters L(model) is the likelihood of the model
- n is number of observation based on the model selection criterion stated above. The model with smallest AIC and BIC value is considered as better fit model but in most cases we cannot find both AIC and BIC minimum. Therefore, in this case we use the most commonly used method of model selection criteria which is called Akaki Information criteria (AIC).

3. Results and Discussion

The aim of this study is to identify determinant factors associated with mortality status of adults and to examine the extent of variations between and within regional states in Ethiopia. The data set used, in this thesis, was the 2011

Ethiopian Demographic and Health Survey conducted under the Ministry of Health and implemented by the Central Statistical Agency and partner organizations, from September 2010 - June 2011. A nationally representative sample of nearly 18500 households was considered. The analysis was performed using data of 55689 individuals. SPSS and STATA software were employed analyses.

3.1. Descriptive Results

The total number of adults covered in this study is 55689 who were in working age group (15-49). Among these, 32.7% reside in urban areas where as 67.3% reside in rural areas. The Chi-square test of association is employed to investigate the association of the potential factors with adult mortality status. Then after, multiple covariates binary logistic regression and multilevel binary logistic regression analysis are applied. The inferential analysis is carried out in two parts. In the first part, determinants of mortality among adults are presented by using binary logistic regression. Finally, we identify factors and variations of mortality of adults across regional states by applying multilevel binary logistic regression model.

3.2. Test of Association Between Adult Mortality and Predictors

From Chi-square Test of association between adult mortality and categorical explanatory variables such as: regions as a dummy variables, age of adults, place of residence, source of drinking water, types of toilet facility, sex of household head, age of household head, types of cooking fuels are found to be significant at 5% level of significance indicating that, association with adult mortality. And household wealth index (quintile) is also significant at 10% level of significant indicating that there is significant association between adult mortality and wealth index. But sex of adults is found to be insignificant at 10% significance level suggesting that no associations with adult mortality.

3.3. Assessment of Goodness Fit of Binary Logistic Regression Analysis

3.3.1. Likelihood Ratio Test

From Table 1, the deviance statistic is the difference between two times likelihood of empty and full model i.e. χ^2 =4113.945 and it is compared with Chi-square at difference of full and empty model degree of freedom (χ^2_{tab} = 33.924 at df =22) as shown in Table 1 Since, χ^2 =4113.945 is greater than χ^2_{tab} = 33.924 with P-value = 0.000, which is less than 0.05 level of significance, indicating an evidence against null hypothesis that, there is no difference between model without adult mortality indicators and with adult mortality indicators. Therefore, including the explanatory variables in the model compared to that of a model with constant term only has better predicting power. So, the conclusion draws based on binary logistic regression analysis including all indicators of adult mortality is appropriate.

Table 1. *Likelihood Ratio Test of Overall Model.*

Model	-2*log likelihood	Df	χ^2	Over all Df	p-value
Empty	18810.809	2	4113.945	22	<0.0001
Full	22924.754	24			

3.3.2. Omnibus Tests of Model Coefficients

From Table 2, the Chi-square value = 234.608 is greater than χ^2tab = 47.400, at df =33 with P-value < 0.0001, which is less than 0.05 level of significance as shown in Table 2. It indicated that, considering all adult mortality indicators together, the overall binary logistic regression model is significantly predicted adult mortality.

Table 2. *Omnibus Tests of Model Coefficients.*

	Chi-square	Df	p-value
Step	234.608	33	<0.0001
Block	234.608	33	<0.0001
Model	234.608	33	<0.0001

3.3.3. Hosmer and Lemeshow Test of Goodness Fit

This test indicates the fitness of predicted and actual value of adult mortality based on EDHS data. The result presented in Table 3 shows the inferential goodness-of-fit test statistics (The Hosmer and Lemeshow Test) that yield a χ^2 vale of 10.699 at 8 degree of freedom and it is insignificant at 5% level of significant, suggesting that the model was fit to the data well. This means that the null hypothesis of a good model fit to data was tenable. In other words the Hosmer-Lemeshow goodness of- fit test statistic is not significant indicating that we do not have an evidence to reject the null hypothesis, suggesting that the model fitted the data well. So, binary logistic regression of adult mortality fitted the EDHS data very well.

Table 3. *Hosmer and Lemeshow Test of Goodness of Fit.*

Chi-square	Df	P value
10.699	8	0.219

3.3.4. Validation of Predicted Probability

The degree to which predicted probabilities agree with actual outcomes is expressed as a classification table. Classification table documents the validity of predicted probabilities. In addition to the measures of association, a classification table documents the validity of predicted probabilities. With the cutoff set at 0.5, the prediction for adults who alive is more accurate than who died. This observation is also supported by the magnitude of sensitivity (15%) as compared to that of specificity (100%). As shown in Table 4, about 100% of adults in Ethiopia lived predicted correctly in binary logistic regression model for adult mortality, while 15% of adult died are predicted correctly in this model. About 96% correct predictions of overall adult is modeled by using binary logistic regression model. The overall correct prediction indicating that there is an improvement over the chance level.

Table 4. *Classification Table.*

Observed		Predicted mortality status of adults		
		Alive	Died	Perc correct
Mortality status	Alive	53438	100	100
	Died	2251	340	15
Overall percent				96

3.4. Multilevel Logistic Regression Analysis

In the multilevel binary logistic regression analysis two-level clusters are used with regions as the second-level units and adults as the first level units (i.e. adults are nested in regions). This study is basically, modeling of adult mortality variations among regional states based on Ethiopia Demographic and Health Survey 2011 data, collected on 55689 adults. The multilevel process was stepwise. The first step examined the null model of overall probability of adult mortality without adjustment for predictors. Second step included both the analysis of single and multilevel model for random intercept and fixed slope multilevel analysis. Third step considered a model for two level random intercept and random slope (Random Coefficient) multilevel logistic regression analysis. The Wald χ^2 test was used to determine significance of each variable of individual β coefficients and log likely hood is used to determine the significance of each model.

3.4.1. Test of Heterogeneity of Adult Mortality Between Regions

The two-level structure is used with the region as the second-level unit and adults as level one unit. This is based on the idea that there may be differences in adult mortality between regions that are not captured by the explanatory variables and hence may be regarded as unexplained variability within the set of all regions. Before attempting to multilevel analysis, we should test the heterogeneity of adult mortality among regional states. As it can be shown in Table 5, the Pearson Chi-square (χ^2_{cal} = 78.110 which is greater than χ^2_{tab} = 18.307. at 10 degree of freedom with P-value = 0.000 which, is less than 0.05 level of significance, implying strong evidence of heterogeneity for adult mortality across regional states. Therefore, multilevel logistic regression model is attempted.

Table 5. *Chi-Square Tests of Heterogeneity of Adult Mortality among Regional States.*

Statistics	Value	Df	p-value
Pearson Chi-square	78.110	10.0	0.000
N of Valid Cases	55689		

3.4.2. Multilevel Empty Model

The single level and multilevel empty (null) models, being compared, differ only with respect to the variance component and it is tested by the deviance of likelihood ratio based on Chi-square. Based on Table 6 the probability of deviance based on Chi-square =43.2054 is greater than χ^2_{tab} = 3.841 at one degree of freedom with P-value =0.000, which is less than 0.05 level of significance. Therefore, multilevel empty binary logistic regression analysis is found to be significant, suggesting that evidence of regional effects on adult mortality in Ethiopia. The overall mean of adult mortality is estimated at -3.161 and the between-region (level two) variance of adult mortality is estimated as σ^2_{u0}=.1727 which is found to be significant at 5% level of significance, indicating the variations of adult mortality among regional states was non-zero. And the intra-region correlation of regional states is estimated at.0099 which is found significant at 5% level of significance, suggesting about 1% of the variation in adult mortality could be attributed to differences across regional states.

Table 6. *Results of Multilevel Empty Binary Logistic Regression Analysis.*

Parameter	Estimate	S.E.	Z	P > Z	95% CI	
					Lower	Upper
Fixed part						
β_0	3.161	.057	-55.78	.000	3.049	3.272
Random part						
σ^2_{u0}	.173	.043	3.987	.000	.106	.282
IR Correlation						
ICC ρ	.009	.005	2.012	.022	.003	.024

Likelihood-ratio test of rho=0: chibar2 (01) = 43.21 Prob>= chibar2 = 0.000

3.4.3. Results of Random Intercept

The random intercept model is where the intercept is allowed to vary across regions after controlling for covariates of adult mortality. It means that, the intercept (β_0) is shared by all regions, while the random effect U_{0j} is specific to region j and the random effect is assumed to follow a normal distribution with variance σ^2_{u0} Here, the random intercept binary logistic regression analysis for adult mortality is found significant based on the difference between log-likelihood of multilevel empty and random intercept binary logistic regression analysis. The probability of deviance based on Chi-square = 79.6411 is greater than χ^2_{tab} = 35.172 at 23 degree of freedom with P-value = 0.000, which is less than 0.05 level of significance. This suggests that, after controlling all indicators of adult mortality, the intercept varied across the region (i.e. the variations of adult mortality among regional states was non-zero).

The variance component for constant term is found significant at 5% level, indicating strong evidence of the variations of adult mortality across regions. The intra-region correlation is found to be 0.0092 implying percentage of the variation of adult mortality could be attributed to the differences between regions. The between region (level two) variance of constant term for adult mortality is estimated as σ^2_{u0} = 0.1744 which is increased by about 0.0166 as compared to empty model. Thus, there is a contribution of those significant factors on adult mortality variations across regional states. The results revealed that, adult's age, source of drinking water, types of toilet facility, household head sex, household head age, types of cooking fuels and household wealth index are found to be significant, indicating strong effects on adult mortality and also contributing to adult mortality variations among regional states. See Table 7. However, place of residence has no significant impact on mortality of adults (p = 0.253, OR=0.9004).

Table 7. *Results of Random Intercept Logistic Model Analysis.*

Variable	Estimates	SE	Z	P>Z	Odds	95% CI Lower	Upper
Adult's Age: Ref: 15-19							
20-24	-.0776	.0767	-1.01	.311	.9253	.7962	1.0753
25-29	-.0246	.0766	-0.32	.748	.9757	.8397	1.1336
30-34	.2966	.0739	4.01	.000*	1.3453	1.1637	1.5553
35-39	.0589	.0844	0.70	.485	1.0607	.8990	1.2514
40-44	.3333	.0869	3.83	.000*	1.3956	1.1768	1.6551
45-49	.2874	.1052	2.73	.006*	1.3330	1.0847	1.6382
Place: Ref: Urban Rural	-.1049	.0918	-1.14	.253	.9004	.7521	1.0779
Water: Ref: Non Improved Improved	-.1048	.0530	-1.98	.048*	.9005	.8116	.9991
Toilet: Ref: Non Improved Improved	-.1397	.0620	-2.25	.024*	.8696	.7701	.9821
HH Sex: Ref:Male Female	.1581	.0492	3.21	.001*	*1.1712*	1.0636	1.2897
HH Age: Ref: 13-24							
25-38	.2763	2.69	.1027	.007*	1.3182	1.0778	1.6122
39-52	.4931	.1048	4.70	.000*	1.6374	1.3332	2.0108
53-65	.3522	.1129	3.12	.002*	1.4222	1.1399	1.7743
65+	.4885	.1255	3.89	.000*	1.6299	1.2745	2.0843
Cooking fuel: Ref: Electricity							
LPG/Ngas/biogas	.2714	.3379	-0.80	.422	.7623	.3931	1.4783
Kerosene	.2393	.2083	1.15	.251	1.2703	.8446	1.9106
Cool/lignite/charcoal	.4242	.1953	2.17	.030*	1.5283	1.0423	2.2409
Wood/crop/grass	1.8585	.3636	5.11	.000*	6.4142	3.1452	13.0808
Other types	.2613	.2019	1.29	.196	1.2986	.8742	1.9293
Wealth index: Ref: Lowest							
Second	.0961	.0754	1.27	.202	1.1009	.9496	1.2763
Middle	-.0558	.0815	-0.68	.493	0.9457	.8062	1.1094
Forth	-.0164	.0773	-0.21	.832	0.9837	.8455	1.1445
Highest	-.2484	.1022	-2.43	.015*	0.7800	.6384	.9530
Constant	3.649	-.2484	14.70	.000*			
Random part							
σ^2_{0u}	.1744	.0445	3.919	.000*		.1057	.2877
ICC ρ	.0092	.0046	2.191	.015*		.0034	.0245

For the age predictor, the trend of odds increases with age of adults. However, only the age groups 30-34, 40-44 and 45-49 are significant at 5% level. The odds The odds of adults with their age group 30-34 is 1.3453 (p = 0.00, CI: 1.1637, 1.5553) which implies that the probability that an adult in this age group dies is increased by 34.53% as compared to the reference age group. The odds of adults with their age group 40-44 is 1.3956 (p = 0.000, CI: 1.1768, 1.6551) which implies that the probability that an adult in this age group dies is increased by 39.56% as compared to the reference age group. The odds of adults with their age group 45-49 is 1.3330 (p = 0.006, CI: 1.0847, 1.6382) which implies that the probability that an adult in this age group dies is increased by 33.30% as compared to the reference age group (15-19) years old.

Source of drinking water is significantly associated with mortality status of adults at 5% level. The odds for adult who uses improved source of drinking water in their household is 0.9004 (p = 0.048, CI: 0.8116, 0.9991) which implies that the probability that an adult who uses improved source of drinking water dies is reduced by 10% as compared to those adults that uses non improved source of drinking water. Type of toilet facility is also significantly associated with mortality status of adults at 5% level. The odds of adults who have improved types of toilet facility in their household is 0.8696 (p = 0.024, CI: 0.7701, 0.9821) which implies that the probability that an adult being died during their productive age is reduced by 13.03% as compared to adults with non improved types of toilet facility in their household. And the odds of adults in which, there is female household head being died during their productive age is 1.1712 (p = 0.001, CI: 1.0636, 1.2897) which implies that the probability of an adult with female household head is increased by 17.12% as compared to adults with male household head.

Likewise, for the age of household head predictor, the trend of odds increases with the age of household head. The odds of adults with age of household head 25 up to 38 and 39 up to 52, are 1.3182 (p = 0.007, CI: 1.077, 1.6122) and 1.6374 (p = 0.000, CI: 1.3332, 2.0108), respectively. This implies that the probability of an adult with household head age 25-38 and 39-52 being died is increases by 31.82% and 63.74% as compared to age group 13-24 years old respectively. While, the odd of adults being died during productive years of age for adults with household head age group 53 up to 65 and above 65, is 1.4222 (p = 0.002, CI: 1.1399, 1.7743) and 1.6299 (p = 0.000, CI: 1.2745, 2.0843), respectively, This implies that the probability of an adult with age of household head between 53 and 65 and above 65 being died is increase by 42.22% and 62.99% respectively as compared to an adult with household head age between 13 and 24.

Type of cooking fuel is significantly associated with mortality status of adults at 5% level. However, only coal/lignight/charcoal and wood/crops/grass/animal dung are significant at 5% level. The likelihood of adults in which coal/lignight/charcoal used mostly as cooking fuels in their household, being died during their productive age is 1.5283 (p = 0.030, CI: 1.0423, 2.2409) which implies the probability that an adult dies with the household uses coal/lignight/charcoal is increases by 52.83% as compared to an adult uses electricity in the household. And the odds of adults in which wood/crops/grass/animal dung mostly used as cooking fuel, being died their productive age is 6.4142 times (p = 0.000, CI: 3.1452, 13.0808) higher than adults who used Electricity in their household. Finally, wealth quintile is significantly associated with mortality status of adults. However, only highest category of wealth index is significant at 5% level. The odds of adults categorized with in higher wealth quintile, being died during their productive age is 0.7800 (p = 0.015, CI: 0.6384, 0.9530) which implies that the probability an adult with the highest wealth quintile being died reduced by 22% as com- pared to adults with in lowest level household wealth index.

3.4.4. Analysis of Random Intercept and Slopes Logistic Model

Models fitted previously have allowed the probability of adult mortality to depend on the region through individual characteristics. This was achieved by allowing the model intercept to vary randomly across regions. We have assumed, however, that the effects of individual characteristics such as cooking fuels and wealth index are the same in each region, i.e. the coefficients of all explanatory variables are fixed across regions. Because it is possible that the effect of wealth index and types of cooking fuels varies across regions these slopes are treated as random. We now ex-tend the random intercept model by allowing both the intercept and the coefficient of wealth index and types of cooking fuels to vary randomly across regions.

The result show that random coefficient estimates for intercepts and the slopes vary significantly at 5% level of significance, which implies that there is a considerable variation in the effects of types of cooking fuels and wealth index, i.e. these variables differ significantly across regions. The random slope binary logistic regression result displayed in Table 8. The variance component for the intercept in the random coefficient model is 0.375951, which is still large relative to its standard error of 0.1655291 with p-value 0.012. Thus there remains some regional-level variance not accounted for in the model. The variance component corresponding to the slope of types of cooking fuels is.0132113, which is also relatively large with respect to its standard error.005706 with p-value 0.01. Thus, this suggests that the effect of types of cooking fuels may be justified in constructing the effect to be random.

The effect of type cooking fuel Lignite/Coal/charcoal as log-odds of adult mortality in region j is estimated as $0.0170629+U_{1j}$ and the between-regions variance in the effect of types of cooking fuels is estimated to be 0.0132113. The negative intercept-types of cooking fuels covariance estimate is -0.0668699 for adult mortality variations across regions.

Table 8. *Parameter Estimates for Random Intercept and Random Slopes Model for Adult Mortality in Ethiopia (EDHS 2011).*

Fixed effects	**Estimate**	**S.E**	**Z**	**P>z**
Adult's Age (ref: 15-19)				
20-24	-.0026233	.002766	-0.95	0.343
25-29	-.0008889	.0027976	-0.32	0.751
30-34	.0120667	.0028933	4.17	0.000*
35-39	.0020204	.0031734	0.64	0.524
40-44	.0141134	.0035535	3.97	0.000*
45-49	.0120152	.0043527	2.76	0.006*
Place Ref: Urban				
Rural	-.005963	.004832	-1.23	0.217
Water Ref: Non Imp. Improved	-.0033509	.0020304	-1.65	0.099
Toilet ref: Not Imp.				
Improved toilet	-.0052461	.0024288	-2.16	0.031*
HH Sex Ref: Male				
Female	.0062702	.0019736	3.18	0.001*
HH Age (ref: 13-24)				
25-38	.0090441	.0034741	2.60	0.009*
39-52	.0174967	.0036192	4.83	0.000*
53-65	.0116129	.0039151	2.97	0.003*
65+	.0169714	.0045832	3.70	0.000*
Cooking fuels ref: Electricity				
LPG/NG/BG	-.0094491	.0114534	-0.83	0.409
Kerosene	.0085687	.0077168	1.11	0.267
Coal/Lignite/charcoal	.0170629	.0071756	2.38	0.017*
Other types	.0112862	.0074098	1.52	0.128
wood/crop./grass	.127455	.0225449	5.65	0.000*
Wealth index ref: Lowest				
Second	.0035357	.0030055	1.18	0.239
Middle	-.0028629	.0033406	-0.86	0.391
Forth	-.001876	.0036546	-0.51	0.608
Highest	-.0108273	.0047837	-2.26	0.024*
Intercept	.0247539	.0103316	2.40	0.017*
Random effects	**Estimate**	**S. E**	**Z**	**P>z**
Variance(U0j)	.375951	.1655291	2.27	0.012
Variance(U1j)	.0132113	.005706	2.32	0.01*
Variance(U2j)	.0019713	.0009202	2.14	0.016*
Covariance(U1j; U2j)	.002078	.0008911	2.33	0.009*
Covariance(U1j; U0j)	-.0668699			
Covariance(U2j; U0j)	-.0181901			
Variance(Residual)	2.902511	.0174016	166.79	0.000*
Deviance-Based Chi-Square	105.40			0.000

*Significant at 5% level; S.E. for Standard Error; Variance(U_{0j}) for intercept; Variance(U_{1j}) for Types of cooking fuels, Variance(U_{2j}) for wealth index.

The random effect of highest wealth quintile on the log-odds of adult mortality in region j is also estimated to be $-0.0108273+U_{2j}$ and the between-regions variance in the effect of wealth quintile is estimated to be 0.0019713. The deviance-based Chi-square value (χ^2 =105.40), shown in Table 8 in appendix, is the difference between the model with and without random effects models. The significance of this difference further indicates that a model with a random slope is more appropriate to explain regional variation than a model with fixed effect coefficients.

The correlation between the intercept and random slope of wealth quintile is -0.0181901. In general, Positive correlation between intercepts and slopes implies that regions with higher intercepts tend to have on average higher slopes and 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 predictor.

3.5. Discussion of Results

This study is aimed to model adult mortality and its variations among regional states based on data from Ethiopia Demographic and Health Survey 2011 (EDHS 2011). Accordingly, different models are fitted to the data to identify potential determinants of adult mortality. First, the ordinary logistic regression model was fitted to the data and significant variables were considered for further investigation in multilevel models. Secondly, the multilevel models were fitted. The intercept only or the empty model is fitted to check whether multilevel effect or heterogeneity exists among the hierarchies. Then the random intercept and fixed slope model is fitted, and finally the random intercept and random slope is fitted. The results are discussed in the following paragraphs.

The variables, having significant association with adult mortality (based on Chi-square test of association) are age of adults, place of residence, region of residence, source of drinking water, types of toilet facility, sex of household head, age of household head, types of cooking fuels used and wealth quintile. The variable sex of adults has no significance association with adult mortality.

Multilevel effects are all significant and have to be taken into consideration in logistic regression modeling. The results from random intercept binary logistic regression model revealed that the overall mean of adult mortality varied across the regions. Analysis of the multilevel random intercept model indicated that, the effects of age of adults, source of drinking water, types of toilet facility, sex of household head, age of household head, types of cooking fuels and wealth index are found to be significant implying significant effects on adult mortality. The impact of place of residence is found to be insignificant at 5% level of significance.

The overall mean for this model of adult mortality is estimated to be -3.65 and between-region (level two) variance of adult mortality is estimated as $\sigma^2_{u0} = 0.174$ which is increased by about 0.00162 as compared to the empty multilevel model. Thus, indicating that, there is a contribution of those factors that have significant effects to adult mortality variations between regional states. The intra-region correlation is estimated at 0.0092, implying that about 1% of the variations in adult mortality is attributed to differences between regions.

Socio-economic status is an important predictor that explaining adult mortality. The odds of adults in the household with highest wealth quintile, being died during their productive years of age are reduced by 22% as compared to those adults with lowest quintile. Generally, mortality of adults goes down as the standard of living of adults rises. This result is consistent with the previous study by [11].

The probability of death for an adult with age group 20 - 24 or 25 - 29 is lower than that in age group 15 - 19 years old. But adults in age group 30 - 34, 35 - 39, 40 - 44 and 45 - 49 are more likely to die as compared to the reference group 15 - 19 years old adults. This result generally shows the trend of adult mortality rate increases with age. This is consistent with reports by [4] and [12]. The probability that, adults live in urban centers die during their productive years of age is not significantly different from those adults live in rural areas. The previous study in China indicated that no change was found in the risk of death or in the urban/rural gap from 1989 through 1997 in country. The result is also consistent with the previous study [12].

4. Conclusions

The purpose of this study is to investigate adult mortality variations among regional states based on the data from Ethiopia Demographic and Health Survey 2011 (EDHS 2011). In this study the classical and multilevel binary logistic regression models are applied to fit a good model of adult mortality on its indicators. This study revealed that socio-economic, demographic and proximate variables have important effect on adult mortality in Ethiopia. Age of adult, sex of household head, age of household head, region of residence, wealth index, types of cooking fuels used and types of toilet facility were the most important determinants for adult mortality in Ethiopia. Place of residence (urban/rural) is not significant factor for the adult mortality.

The result also showed that the random intercept binary logistic regression model best fits to the data. The effect of regional variation for wealth index and types of cooking fuel further implies that there exists considerable difference in adult mortality among regions. The results from the random intercept multilevel logistic regression analysis suggest that there exist significant differences in adult mortality among regions in Ethiopia. A model with random effects is more appropriate to explain the regional variation than a model with fixed effects.

References

[1] World Health Organization WHO (2014). World Health Statistics 2014. Geneva.

[2] Richard G., Hummer, A. and Patrick, A. (2003). Adult Mortality. Working Paper POP 2003, 0003.

[3] Chasimpha, S., McLean, E, Chihana, M, Kachiwanda, L, Koole1, O, Tafatatha, T, Mvula, H, Nyirenda, M, Crampin, AC, Glynn, JR (2015). Patterns and Risk Factors for Deaths from External Causes in Rural Malawi over 10 Years: A Prospective Population-based Study. BMC Public Health, 15:1036. DOI: 10.1186/s12889-015-2323-z.

[4] CSA (2011). Ethiopia Demographic and Health Survey Report.

[5] Hosmer D and Lemeshow S (2000). Applied Logistic Regression. Second Edition. New York: John Wiley and Sons.

[6] Agresti, A. (1996), Categorical Data Analysis. Second edition. University of Florida, John Wiley & sons, INC.,Publication

[7] Bewick L. and Jonathan B. (2005). Statistics Review 14: Logistic Regression.

[8] Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage.

[9] Connel A. A. (2006). Logistic Regression Models for Ordinal Response Variables. Thousand Oaks: Sage. QASS No. 146.

[10] Taper M. L. (2004). Model Identification from Many Candidates. The Nature of Scientific Evidence. M. L. Taper and S. R. Lele, Editors. The University of Chicago Press, Chicago, IL, USA. P 3-16.

[11] Saikia, N and Bhat, P. N. (2008), Factors Affecting Adult Mortality in India: An Analysis of National Family Health Surveys of 1992-1993 and 1998-99. Journal of Demography India, Vol. 37.

[12] Zimmer, Z., Kaneda, T, Spress, L (2006). Urban Versus Rural Mortality among Older Adults in China. Working paper, New York, New York 10017 USA.