Jimma University

College of Natural science

Department of statistics

Statistical Analysis of Anemic Status among Pregnant Women in Ethiopia

By:

Assaye Belay

A Thesis submitted to the department of Statistics, School of Graduate Studies, College of Natural Science, and Jimma University in Partial Fulfillment for the Requirements of Master of Science (M.Sc.) in Biostatistics

December, 2014

Jimma, Ethiopia

Statistical Analysis of Anemic Status among Pregnant Women in Ethiopia

By:

Assaye Belay

Advisor: Wondosen Kasshun (Assistant professor, PhD)

Co-advisor: Kibrealem Sisay (M.Sc)

December, 2014

Jimma, Ethiopia

Department of Statistics, School of Graduate Studies

Jimma University

As thesis research advisors, we herby certify that we have read and evaluated the thesis prepared by **Assaye Belay Gelaw** under our guidance, which is entitled Statistical Analysis of anemic status among pregnant women in Ethiopia **.**We recommend that the thesis be submitted as it fulfills the requirements for the degree of Master of Science.

Wondwosen Kasshun (PhD) ---------- ----------

Main advisor Signature date

Kibrealem Sisay (M.Sc) ---------- ----------

Co-advisor signature date

As the members of the board of examiners of MSc. thesis, open defense examination of **Assaye Belay Gelaw**, we certify that we have read and evaluated the thesis and examined the candidate. We recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

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

Name of chairperson Signature Date

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

Name of Main Advisor Signature Date

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

Name of Co-advisor Signature Date

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

Name of internal Examiner Signature Date

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

Name of External Examiner Signature

Statement of Author

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

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

Assaye Belay

Date: _____

Signature: _____

Jimma University, Jimma

Dedication

This thesis is dedicated to my beloved all families, especially to my mother Almtsehay Kebede and my father Belay Gelaw

Assaye Belay

Date: _____

Signature: _____

Jimma University, Jimma

Acknowledgment

First, my heartfelt thanks goes to almighty GOD who helped me to prepare this paper

Secondly, my heartfelt thanks goes to my advisor Wondwosen Kasshun (PhD) for his valuable advice in preparing this research paper.

Thirdly, my heartfelt thanks goes to my Co. Advisor Kibrealem Sisay (MSC) for his cooperation in preparing this research paper

Fourthly, I would like to thank Jimma University School of Graduate studies for its financial support to conduct this study preparation.

Lastly, I would like to extend thanks to all of my friends and family for their support in my academic life

Abstract

Background**:** Anemia in pregnancy is related to different socio-demographic, dietary and economic factors, Mother's age, educational status, economic position, and antenatal care were significantly associated with anemia during pregnancy.

Objective**:** To fit an appropriate statistical model and identify potential factors of anemic status among pregnant women in Ethiopia

Methods**:** A cross-sectional but cluster study carried out based on the secondary data of the Ethiopia Demographic Health Survey. Data of a total of 1277 pregnant women of reproductive age (15-49) were included in the analysis. Data were mainly analyzed using that SAS software offers for the analysis of binary responses with correlated data (GENMOD procedure), and both marginal and cluster specific data (NLMIXED procedure). It has been showed that each of them estimates parameters from among different statistical models and comment on the interpretation of parameters and the statistical properties of the methods involved. For the categorized response variable (non normal response), General linear model, over dispersion, Generalized Estimating Equation, Generalized Linear Mixed model, and Marginalized Multilevel model were used to identify the associated factors of anemic status among pregnant women.

Result: some of covariates for the marginal model revealed that pregnant women those lived in urban had 0.862 (*p = 0.0012*) times lower risk than those who lived in rural or the probability that the pregnant women those who lived in urban had 13.8% times less likely to develop anemia than those who lived in rural, on the other hand the pregnant women whose education status was poor had 2.087(p-value=0.0001) times higher risk to develop anemia than those whose education status was higher .Similarly, the number of pregnant women who had HIV –positive had 1.39 *(p = 0.0001)* times higher risk than their counterparts and similar results were obtained in cluster specific and marginalized multilevel model.

Recommendation**:** Government should design strategies and policies to enhance women education to make them independent in socio-economic and cultural decision, which directly and indirectly affect women health status due to anemia. It is recommended that the remaining factors that have not been included in this study could be included in future studies.

Table of contents

Content page

List of Tables

Table page

List of Acronyms

AIC:	Akaike Information Criteria
AR(1)	Autoregressive order one
CDCP	Center of Disease Control and prevention
CSA	Central Statistics Agency
EDHS	Ethiopian Demographic Health Survey
(EHNRI)	Ethiopian Health and Nutritional Research institute
g/dl	gram per deciliter
GEE	Generalized Estimate Equations
GLM	Generalized linear model
GLMM	Generalized Linear Mixed Model
Hb	Hemoglobin
MMM	Margin zed multilevel model
NRERC	National Research Ethics Review Committee
LBW	Low Birth Weight
GLM	Generalized linear model
ML:	Maximum Likelihood
NLMM	Non linear mixed model
SNNP	South Nation and Nationality of People of Ethiopia
REML:	Restricted Maximum Likelihood
OR	Odd Ratio
WHO	World Health Organization

1.Introduction

1.1. Background of the study

Anemia is one of the most widespread public health problems, especially in developing countries. It impairs cognitive development, reduced physical work capacity and in severe cases increased risk of mortality particularly during prenatal period (WHO, 2001). Anemia in pregnant women is defined hemoglobin levels less than 11g/dl (WHO, 1996). It is usually caused by iron deficiency, which is the most common nutrient deficiency in the world. It has been estimated that, at any one time in developing countries, half of the population (mainly children and women of reproductive age) is affected by anemia (Hercberg and Galan, 1992). During pregnancy, approximately 75% all anemia diagnosed are due to iron deficiency (Sifakis and Pharmacies, 2000). Estimates in Kenya and Nepal suggest that hookworm infection causes 30 percent and 41 percent, respectively, of moderate or severe cases of anemia among pregnant women (hemoglobin level, < 9 g per deciliter. Studies in Africa and Asia reported a higher prevalence of anemia and its association with women of age < 20 years, third trimester of pregnancy, rural residents and multifarious women (Singh and Fong, 1998). Anemia in pregnancy is also related to different socio-demographic, dietary and economic factors. Mother's age < 20 years, educational status, economic position, and antenatal care were significantly associated with anemia during pregnancy in a study conducted in India (Bechuram *et al*., 2006).In Ethiopia, anemia is the most frequent morbidity among pregnant women with the prevalence raging from 23 to 66.5% (Tadios, 1996, Gebremedin, 2004). There is an urban rural difference in the prevalence of anemia.

Anemia is a condition characterized by a decrease in the concentration of hemoglobin in the blood. Hemoglobin is necessary for transporting oxygen to tissues and organs in the body. The reduction in oxygen available to organs and tissues when hemoglobin levels are low is responsible for many of the symptoms experienced by anemic persons. The consequences of anemia include general body weakness, frequent tiredness, and lowered resistance to disease. Anemia can be particularly serious problem for pregnant women, leading to premature delivery and low birth weight. Overall, morbidity and mortality risks increase for individuals suffering from anemia. Anemia is classified as mild, moderate or severe based on the concentrations of hemoglobin in the blood. The cutoffs values used in defining each of these levels vary according to age and, for ever married women, pregnancy status (WHO, 2001).

1.2. Statement of the Problem

Anemia in pregnancy causes significant maternal morbidity and mortality in the developing countries including Ethiopia. The burden and underlying factors are varied even within countries in such a way that anemia in pregnant women is related to different socio-demographic, dietary and economic factors. According to WHO's estimate, the global prevalence of anemia in pregnant women is 68%. In Africa its prevalence is estimated to be 66.8%. In Ethiopia, anemia is the severe public health problem affecting 62.7% of pregnant mothers and 52.3% non-pregnant women. This study, therefore, attempts to identify determinant factors of the case of anemia among pregnant women in Ethiopia and considering clustered data from EDHS, 2011 by addressing the following research questions:

- ✓ Which model is best fit for the data of anemic among pregnant women?
- ✓ Which covariates are the most determinant factors for anemic among pregnant women?
- ✓ Is there a significant variation or regional heterogeneity as well as cluster heterogeneity in suffering from anemic among pregnant women?

1.3. Objective of the Study

1.3.1. General Objective

To assess the magnitude and factors associated in it's anemic in pregnant women

1.3.2. Specific Objectives

The specific objectives of the study which should be accomplished to achieve the general objective stated above are:

- ✓ To estimate the prevalence of anemia among pregnant women in Ethiopia.
- ✓ To fit an appropriate statistical model and identify the potential factors affecting anemic status among pregnant women in Ethiopia.
- ✓ To see the cluster variation of anemia among pregnant women

1.4. Significance of the Study

This study is mainly useful to understand how the cluster structure of the data in which the magnitude of the random effects and correlation structure are under consideration in the analysis. It may serves as stepping-stone for those who are interested to undertake an in depth research on issues related to the death of pregnant women due to anemia in Ethiopia. This study will help to the stakeholders to reduce maternal and infant mortality rate due to the severity of anemia, and clarifying the main determinant factors that significantly affecting among pregnant women due to anemia. Generally, this research is expected to give idea to those focuses on this area:

- The results of this study will give information to concerned bodies in setting, policies, strategies and further investigation for decreasing the severity of anemia among pregnant women in Ethiopia
- To give emphases on the factors those have strong association with pregnant women so that policy makers act on accordingly.
- To introduce(familiarize) different statistical model for analyzing biological as well as socio- demographic factors for health staffs as well as related researchers
- I hope, this study may also be used as a stepping-stone for further studies.

2. Literature Review

2.1. Burden of anemia in pregnancy

The health conscious world community has come to realize that anemia, the majority of which is due to iron deficiency, has serious health and functional consequences are wide spread especially among tropical low income populations and that most of its nutritional component is controllable with a very high benefit/cost ratio. Women of fertile age and pregnant–lactating as well as their infants and young children are particularly affected (WHO, 1991).

Anemia in pregnancy, (hemoglobin level <11g/dl as defined by World Health Organization is a major public health problem, especially in developing countries (de Benoist *et al*., 2008) . Recent statistics indicate that anemia affects 41.8% of pregnant women globally, with the highest prevalence in Africa (WHO, 2006). Fifty seven percent of pregnant women in Africa are anemic, which corresponds to about 17 million affected women, with severe consequence on health, social, and economic development (de Benoist *et al*., 2008). Studies in Africa have shown a high prevalence of anemia in pregnancy ranging from 41 to 83% in different settings (Meda *et al*., 1999). There is however significant variation in prevalence of anemia, both within and between countries, necessitating a need for local data to help inform preventive programs. Anemia in pregnancy is associated with negative consequence for both the woman and neonate. Foetal anemia, low birth weight (LBW), preterm birth, intrauterine growth restriction, and prenatal mortality have been associated with anemia (Scholl, 1994; Kidanto *et al*., 2009). In the women themselves it may cause low physical activity and increased risk of maternal morbidity and mortality, especially in those with severe anemia (Scholl, 1994; Allen, 2000; de Benoist *et al*., 2008)

2.2. Several Risk Factor for Anemia among Pregnant Women

Poor dietary status reflected by low socio- economic status makes micronutrient deficiency both clinical and subclinical relatively more common. All these factors deplete the micronutrient stores of the mother, to the extent that she becomes anemic and this brings a more severe outcome for both the mother and the child reported by (Bondevik &Abel, 2001). Yuan Xing et al (2009), on study reported that an average of 63 percent of mothers were anemic and that the gestational age, ethnicity, residence and low income amounted significantly to the Hb level and the occurrence of anemia in pregnant mothers and reported 41.58% in pregnant women of Qingdao province of China were anemic and the subjects with iron deficiency anemia had much higher rates of vitamin C, foliate and

B12 deficiencies than those in the non anemic subjects and especially in the deficient rates of ascorbic acid and foliate in the anemia group. Moreover, they observed that the decreasing trends of Hb concentrations were accompanied by the decreases of serum levels of vitamin A, ascorbic acid, foliate and B_{12} and concluded that multiple vitamin deficiencies may be associated with anemia in pregnant mothers in the last trimester.

However, the work of Karaoglu et al (2010) on pregnant women of East Anatolian province of Turkey, reported percentage of 27.1% of anemic pregnant women, having four or more children and being in the third trimester. Their finding also was associated with soil eating habits of pregnant women. Most of the anemia recorded was norm ocyticnormochromic indicating mixed anemia.

Many studies explained the status of anemia in pregnant mothers depended on the socioeconomic level (Idowu et al (2005),illiteracy, extremes of mother's age, grand gravida, short pregnancy intervals and age of gestation .In measuring the status of anemia in the population, hemoglobin (Hb) concentration is the most reliable indicator as opposed to clinical measures.

3. Methodology

3.1. Source of Data

The source of data for this study has been used the 2011 Ethiopia Demographic and Health Survey (EDHS), which obtained from Central Statistical Agency (CSA). It was the third survey conducted in Ethiopia as part of the worldwide Demographic and Health Surveys project. The 2011 Ethiopian Demographic and Health Survey, was designed to provide estimates for the health and demographic variables of interest for the following domains. Ethiopia as a whole; urban and rural areas (each as a separate domain); and 11 geographic administrative regions (9 regions and 2 city administrations), namely: Tigre, Afar, Amharic, Oromiya, Somali, Benishangul-Gumuz, Southern Nations, Nationalities and Peoples (SNNP), Gambela and Harare regional states and two city administrations, that is, Addis Ababa and Dire Dawa.

The principal objective of the 2011 EDHS was to provide current and reliable data on fertility and family planning behavior, child mortality, adult and maternal mortality, children's nutritional status, use of maternal and child health services, knowledge of HIV/AIDS, and prevalence of HIV/AIDS and Anemia in general and the sample size of the population under the study was 1277 pregnant women.

3.2. Variables

3.2.1. The dependent variable

Often in many epidemiologic, biomedical and related fields of studies, the outcome of interest is a binary variable such as anemic versus non anemic. In such circumstances, it is possible to employ plausible statistical tools for estimating the magnitude of the association between the response variable of interest as a function of independent predictor variables. The association provides information about the risk of developing an outcome. In practical advantage of using statistical methods for binary response over statistical methods for continuous response variable in epidemiologic research is that parameter estimates of the possible risk factors can be directly converted to an odds ratio, which is interpretable. Additionally, the use of binary outcome for defining anaemia and its severity at the population level, as well as the chronology of their founding allows the identification of populations at greatest risk of anaemia and priority areas for action, especially when resources are inadequate. In view of the above, the hemoglobin level was first dichotomized based on the cut-off points as described in literature view leading to the binary response:

$respnse(anemic\ status) = \begin{cases} 1\ if\ Hb\ level\ < 11 g/dl \\ 0\ if\ Hb\ level\ \geq 11\ g/dl \end{cases}$ Where 1 was coded for anemic and 0 has coded as non anemic.

3.2.2. Predictor (explanatory) variables

The explanatory variables that would be included are explained as follow. The choice of these variables is guided by different literatures as the determinant factors for anemia among pregnant women. These categories of the independent variables were coded starting from zero to make it appropriate for further analysis using different statistical models. Such explanatory variable are : Age, Region, Religion, Residence, occupation, smoking status, HIV+/-, wealth index, Vitamin intake, marital status and education.

Table 3.1 Coding and description of explanatory variables

Covariates	Description	categories
age	age of individual	(15-26)=0,(27-38)=1,(39-50)=2
estatus	education status	0=no edu,1=primary,2++=secoundary,3=higher
occup	occupation	0=non employed,1=employed
Rel	Religion	1=ortho 2=catho 3 =prot,4=musl,5=trad
winx	Wealth index	1=poorest,2=poor ,3=middle.4=riche,5=richest
sm.status	smoking status	0=non smoked,1=smoked
mar. Status	marital status	0=sing,1=marr,2=llwp,3=widow,4=divor,5=nollwp
vitk	vitamin intake	0=no vitamin intake,1=vitamin intake
resid	residence	2=rural ,1=urban
HIV	HIV	0=HIV-,1=HIV+
Reg	Region	1=tgray,2=Afar,3=Amhara,4=Oromiya
		5=Somali,6=B.Gumz,7=SNNP,8=Gambela
		9=Harar,10=AddisAbaba,11=Diredawa

Where nollwp =no longer lived with partner, llwp=long lived with parner

3.3. Statistical Models

3.3.1. Generalized Linear Model

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

Generalized linear model assumes the response variables are independent (the dependent variable which is anemic status are independent or uncorrelated. In clustered data, observations are usually taken from the same unit, and thus this information forms a cluster of correlated observations. For instance, in the EDHS the dependent variable (anemic status for pregnant women) was measured once for each eligible mothers nested within clusters from each region.

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

A random variable Y follows a distribution that belongs to the exponential family, if the density function is of the form

$$f(y/\theta,\phi) = exp\{\phi^{-1}[y\theta - \psi(\theta)] + c(y,\phi)\}$$

, for a specific set of unknown parameters θ and ϕ, and for known functions ψ(·) and c(·,·). The parameter θ is called the canonical parameter and represents the location while, ϕ is called the dispersion parameter and represents the scale parameter and for the Poisson and binomial

distribution it is fixed to be one (Faraway, 2006). An important property of the GLM is the functional relation between mean and variance. .

3.3.2. Over dispersion Model

In practice, many types of outcomes using standard models within the GLMs for their analysis, such as binomial and count observations, often exhibit variability exceeding what is predicted by the binomial or Poisson distribution (Molenberghs et al., 2010).The standard Bernoulli model assumes that the mean and variance depend on single parameter. Though a set of i.i.d. Bernoulli data cannot contradict the mean variance relationship, it may not hold true for data having a hierarchical structure of the form z_i successes out of n_i trials, such as in cluster and longitudinal studies.

A simple quasi-likelihood approach uses the variance function$var(\pi_i) = \emptyset * \frac{\pi_{i(1-\pi_i)}}{n_i}$. In this context, if $\emptyset > 1$, over dispersion is said to occur. An elegant way to account for over dispersion in clustered binary and binomial data is through inclusion of beta random-effects, leading to the so-called beta-binomial model, in which the Bernoulli model is combined with a beta Distribution (Molenberghs and Verbeke, 2005).

3.3.2.1. Beta –Binomial model

The beta-binomial model can be introduced by requiring conjugacy on the one hand or, as done here, it can be generated from first principles (Skellam 1948, Kleinman 1973, Molenberghs and Verbeke 2005) on the other. The model follows from mixing the binomial parameter over a beta distribution. Suppose that $\frac{z_i}{\pi_i} \sim Bin(n_i, \pi_{i)\ and}\ \pi_i \sim beta\ (\alpha_i, \beta_i) where\ 0 \leq \pi_i \leq 1\ with \alpha_i \geq 0\ and$ and $\beta_i \geq 0$.The density, mean, and variance for π_i then easily follow:

$$f_{(\pi_i)=\frac{1}{B(\alpha_i, \beta_i)}}\ (1-\pi_i)^{\beta_{i-1}}\quad \pi_i^{\alpha_{i-1}}$$

$$E(\pi_i) = \frac{\alpha_i}{\beta_i + \alpha_i}$$

$$\text{v}var(\pi_i) = \frac{\beta_{i\alpha_i}}{(\alpha_i+\beta_i+1)(\alpha_i+\beta_i)2}$$

$$f(Z_i) = \int_0^1 f(z_{\frac{i}{\pi_{i)}}} f(\pi_i) d\pi_i = \frac{n_{i!}}{(n_i - z_i)!\, z_{i!}} \frac{\gamma(\alpha_i + \beta_i)\gamma(\alpha_i + \beta_i - z_i)\gamma(\alpha_i + \beta_i)}{(n_i + \alpha_i + \beta_i)\gamma(\alpha_i)\gamma(\beta_i)}$$

Likewise, these elements for $Z_i$are:

$$(E(Z_i) = E(E\left(\frac{Z_i}{\pi_i}\right)) = E(n_i Z_i) = E(Z_i) = n_i \frac{(\alpha_i)}{(\beta_i + \alpha_i)} = n_i \mu_i$$

$$var(z_i) = E(var\left(\frac{z_i}{\pi_i}\right) + var(E\left(\frac{Z_i}{\pi_i}\right) = n_i \mu_i (1 - \mu_i)(1 + (1 - n_i)(\frac{1}{(\beta_i + \alpha_i + 1)}$$

It is easy to show that the correlation between any two outcomes y_{ij} and y_{ik}, $j_{\neq k}$from the same cluster i equals$\rho_i = (\frac{1}{(\alpha_i+\beta_i+1)})$. By using this expression in combination with $\mu i = \frac{\alpha_i}{(\alpha_i+\beta_i)}$, the marginal density can be rewritten as

$$f(Z_i) = \binom{n_i}{Z_I} B\left[\mu_i\left(\frac{1}{\rho_i} - 1\right) + z_i(1 - z_i)\left(\frac{1}{\rho_i}\right) + (n_i - z_i)\right]$$
$$/B(\left[\left[\mu_i\left(\left(\frac{1}{\rho_i}\right) - 1\right), (1 - \mu_i)(\frac{1}{\rho_i}\right]\right] \ldots . (3.4)$$

In applying the beta-binomial model it is common, but not absolutely necessary, $\alpha_i$to assume and β_i constant across i. The parameter is the dispersion parameter which is constrained to be positive in the beta-binomial model. When ρ = 0, the ordinary binomial variance results. Also, for n_i = 1, the Bernoulli model is recovered. Over dispersion occurs when ρ > 0. The beta-binomial model allows for modeling the μ_i's with a linear predictor through a link function$g(\mu_i) = x_{i\beta}^{'}$.The cluster-specific dispersion parameter ρ_i can also be modeled through Fisher's z transformation (Molenberghs and Verbeke 2005).The beta-binomial model assumes the $b_i$to come from a beta distribution with parametersα andβ .The parameter α and βcan depend on covariates, but this dependence is temporarily dropped from notation.

3.3.3. Marginal Model

Marginal models are among the most statistical models widely used to model clustered as well as repeated data. In marginal models, the main scientific objective is to analyze the population-averaged effects of the given factors in the study on the binary response variable of interest. This means that the covariates are directly related to the marginal expectations. The marginal models fitted in this cluster data included the Generalized Estimating Equations (GEE).

For binary data, recently, Balemi and Lee (1999) obtained –finite expansion bias and efficiency of the estimates from GEE approaches with miss-specified correlation matrices. The main

findings are :i) bias and efficiency depend on the combination of a number of characteristics of the data :cluster size, intra cluster correlation covariates , intra cluster correlation response variable ,variability of cluster size and the relative response association and ii)the performance of GEE is excellent for moderate degree of response correlation small clusters.

Furthermore, GEE is non-likelihood method that uses correlation to capture the association within the clusters or subjects in terms of marginal correlations (Molenberghs & Verbeke, 2005). For clustered as well as repeated data, (Liang & Zeger 1986) proposed GEE which require only the correct specification of the univariate marginal distributions provided one is willing to adopt "working" assumptions about the correlation structure. The "working" assumptions as proposed by Liang and Zeger included independence, unstructured, exchangeable and auto-regressive AR (1). Independence and exchangeable working assumptions can be used in virtually all applications, whether longitudinal, clustered, multivariate, or otherwise correlated. Auto regressive AR (1) and unstructured correlation structures are less relevant for clustered data, studies with unequally spaced measurements and/or sequences with differing lengths (Molenberghs and Verbeke, 2005).

3.3.3.1. Independence Structure

In GEE, the model assumes independeent correlation by default. With this structure the correlations between subsequent measurements are assumed to be zero or measurements are independent to each other within individuals in the given cluster .

3.3.3.2. Autoregressive (AR)

Box *et al*. (1994) described the family of correlation structure, which includes different classes of linear stationary models: autoregressive models, moving average models, and mixture of autoregressive-moving average models. Autoregressive models express the current observation as a linear function of previous observation plus a homoscedasticty noise term.

3.3.3.3. Exchangeable correlation structure (compound symmetry)

It assumes the correlations between subsequent measurements are assumed to be the same, irrespective of the cluster data. Generally, assuming no missing data, the J x J covariance matrix y is modeled as $V_i = \Phi A_i^{1/2} R_i A_i^{1/2}$

Where Φ is a glm dispersion parameter which is assumed 1 for binary categorical data, A_i is a diagonal matrix of variance functions, and R_i is the working correlation matrix of Y. Generalized estimating equations (GEEs) can be used to model correlated data with the variance covariance matrix V by iteratively solving the quasi- score equations. The score function of a GEE for β has the form

$$\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \beta_i} \right) V_i^{-1} (Y_i - \mu_i) = 0$$

Where μ_i is the fitted mean, which is given by $g(\mu_{it}) = X_{it}\beta$ for covariates $X = x_{i1}, x_{i2}, \ldots\ldots\ldots, x_{im}$ and regression parameters $\beta = \beta_1, \beta_2, \ldots\ldots\ldots\ldots, \beta_P$ Starting R_i as the identity matrix and Φ=1, the parameters β are estimated by solving equations as follows.

i.e. in normal case $\mu_i = X_i\beta$ and $\frac{\partial \mu_i}{\partial \beta_j} = x_i$, $V_i = \hat{\Phi} R_i$

$$\sum_{i=1}^{N} \left(x_i^{\,t} \right) R_i^{-1} (Y_i - \mu_i) = 0$$

More generally, because solution only depends on the mean and variance of y, these are quasi likelihood estimates. The estimates from a GEE analysis are robust to miss-specification of the covariance matrix (Liang & Zeger, 1986), so, the regression parameter estimates are consistent even for independent covariance matrix. Upon convergence, in order to perform hypothesis tests and construct confidence intervals, it is of interest to obtain standard errors associated with the estimated regression coefficients. These standard errors are obtained as the square root of the diagonal elements of the matrix var($\hat{\beta}$) .

Two models are compared using generalized Wald test for GEE and likelihood ratio test for GLMM or simply backward selection technique.

Let $Y_j = (y_{j1}, \ldots\ldots y_{jn_j})'$ be the response values of observations from j^{th} cluster, $j = 1,2,\ldots,m$ follows a binomial distribution i.e $Y_j \sim Bino(n_j, \pi_j)$ that belongs to the exponential family with the density function of the form. Then to model the relation between the response and covariates, one can use a regression model similar to the generalized linear models given by:

$$g = logit(\pi_j) = X'_j\beta$$

Where, $g(\pi_j)$= logit link function,

X_j= (n_j x p) dimensional vector of known covariates.

β= (1 x p) dimensional vector of unknown fixed regression parameter to be estimated

$E(Y_j) = \pi_j\ is$ expected value of the response variable.

3.3.3.4. Parameter Estimation for GEE

Here GEE is not likelihood approach, rather it is quasi-likelihood based and estimates $\hat{\beta}$ by solving estimating equations which consist of the working covariance matrix V_j. The score equation used to estimate the marginal regression parameters while accounting for the correlation structure is given by: $S(\beta) = \sum_{j=1}^{m} \frac{\partial \pi_j}{\partial \beta'} \left[A_j^{1/2} R_j A_j^{1/2}\right]^{-1} (Y_j - \pi_j) = 0$ .Where R_j is working correlation matrix, and the covariance matrix of Y_j is decomposed in to $A_j^{1/2} R_j A_j^{1/2}$ with A_j the matrix with the marginal variances on the main diagonal and zeros elsewhere, $\boldsymbol{Y_j}$ multivariate vector of asymptotically normal response variables with mean vector π_j i.e $Y_j \cong N(X_j\beta, V_j)$. An advantage of the GEE approach is that it yields a consistent estimator of $\hat{\beta}$, even when the working correlation matrix R_j is misspecified. However, severe misspecification of working correlation may seriously affect the efficiency of the GEE estimators (Molenberghs & Verbeke, 2005).

3.3.4. Generalized Linear Mixed Model (GLMM)

An alternative way to fit a longitudinal (cluster) model to non-normal response data is to fit a generalized linear mixed model. These models are similar to the ones fit in GEE because the normality assumption regarding the error terms is relaxed. Some of the error distributions supported by generalized linear mixed models include the binomial, Poisson, gamma e.t.c. These models also support a large variety of link functions, which include the logit, log, and reciprocal. The type of response variable determines the distribution and link function for the model. Since the response variable for GLMM was categorical, binary data the logit link function was used to identify the associated factors of anemia among pregnant women. However, unlike the models fit in GEE, generalized linear mixed models have the

flexibility to specify random effects and also to generate subject-specific parameter estimates (Verbeke & Molenberghs ,2005). Let Y_{ij} denote the response for i^{t} individual at jth cluster. Y_{ij} is categorical response variable with each follows a binomial distribution.

Generalized linear models (GLMs) is one parts of subject specific models which extends ordinary regression by allowing non-normal responses and a link function of the mean. The generalized linear mixed model is a further extension that permits random effects as well as fixed effects in the linear predictor (Agresti, 2002).

Let y_{ij} denote the response of i^{th}individual pregnant women from j^{th}cluster where $i = 1,2 \dots \dots \dots \dots n_j$ and Y_j the n_j dimensional vector of all measurements available for cluster $j(region)$. Let $f(b_j/D)$ be the density of the $N(0,D)$ distribution for the random effects b_j. Assumed conditionally on q-dimensional random effects b_j to be drawn independently from $N(0,D)$, the outcomes $\boldsymbol{y_{ij}}$ of Y_j are independent with the density of the form

$$f_j(y_{ij}/b_j,\beta,\phi) = exp\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij},\phi\},$$

Then the generalized linear mixed model (Molenberghs and Verbeke, 2005); with logit link is defined $\boldsymbol{as}\ logit(\pi_{ij}) = X'_{ij}\beta + Z'_{ij}b_j,\quad j = 1,2,\dots,m$

Where, $E(Y_{ij}/b_j) = \pi_{ij}$, is the mean response vector conditional random-effects models can be fitted by maximization of the marginal likelihood, obtained by integrating out the random effects. Such likelihood may involve high-dimensional integrals that cannot be evaluated analytically. The likelihood of the data expressed as a function of unknown parameters is on the random effects b_j, for pregnant women in cluster j and, X_{ij} and Z_{ij} are p-dimensional and q-dimensional vectors of known covariate values. The random effects b_j are assumed to follow a normal distribution with mean **0** and covariance matrix *D.*

3.3.4.1. Parameter Estimation for GLMM

$$\mathrm{L}(\beta,D,\phi) = \prod_{j=1}^{m} f_j(Y_j/\beta,D,\phi) = \prod_{j=1}^{m}\int \prod_{j=1}^{n_j} f_{ij}(Y_{ij}/b_j,D,\phi) f(b_j/D)db_j$$

It is the integral over the unobserved random effects of the joint distribution of the data and random effects. The problem in maximizing is the presence of m integrals over the q-

dimensional random effects b_j. With Gaussian data, the integral has a closed form solution and relatively simple methods exist for maximizing the likelihood or restricted likelihood. With non-linear models, numerical techniques are needed.

3.3.4.2. Model Building for GLMM

Under the GLMM for clustering, random effects are included the model to address the between-region and within-region variations. These will be introduced in the generalized linear mixed model due to the fact that, the probability of pregnant women with the severity of anemia are possibly varies for individuals within the same regions as well as individuals in different regions.

3.3.4.3. Model Comparison in GLMM

This study will be used Likelihood ratio test and Information criteria to select the best model based on the values of asymptotic estimations. In order to decide on the best of the two random effects models, two models will be fitted, one with the two random intercepts (between and within clusters variations) and another one with one random intercept(within cluster variation). One can use the approximate restricted maximum likelihood ratio test (LRT) to compare these two models (Myers et al., 2010).

Let $LR_{full} = -2\log likelihood$ value for the full model and $LR_{redu} = -2loglikelihood$ value for reduced model. Then, the likelihood ratio test statistic, is given by

$$\lambda = LR_{full} - LR_{redu}$$

The asymptotic null distribution of the likelihood ratio test statistic λ, is a chi-square distribution with degrees of freedom equal to the difference between the numbers of parameters in the two models.

Akaike's information criterion (AIC)

AIC is a measure of goodness of fit of an estimated statistical model. It is not a test on the model in the sense of hypothesis testing; rather it is a tool for model selection. The AIC penalizes the likelihood by the number of covariance parameters in the model, therefore $AIC = -2\,log(L) + 2p$ .Where, L is the maximized value likelihood function for the estimated model and p is the number of parameters in the model. The model with the lowest AIC value is preferable.

3.3.4.4. Model Checking Technique

In GLMM, it is assumed that the random effects are normally distributed and uncorrelated with the error term. Normality of the random effects is assessed using normal plot of each random effect. Normal Q-Q plot of estimated random effects is an important method for checking the marginal normality and identify outlier.

3.3.5. Marginalized Multilevel Models

Random effects models (such as GLMMs) are applied to longitudinal (cluster) data by specifying a mean model that is conditioned on a set of latent 'random' effects. The latent effects are conceived as embodied sources from which the within- subjects associations arise. GLMMs have many advantages, including the ability to work within a likelihood framework, having subject specific regression coefficients, flexibility in specifying within-subject dependence mechanisms, and valid inferences under missing at random (MAR) dropout mechanisms. Drawbacks of GLMMs include sensitivity of regression coefficients to association structure assumptions and, in many problems, regression parameter interpretations being conditional on unobservable effects (Diggle *et al*. 2002).

Marginalized multilevel models embrace the interpretation and robustness of regression coefficients from a marginal model, while retaining the likelihood inference capabilities and flexible dependence specifications from a GLMM. The MMM formulation given in Heagerty and Zeger (2000) uses a standard GLM for the marginal mean, a non-linear mixed model (NLMM) for the within-subject associations and a specified probability distribution for the underlying latent effects:

$$i)\ g(\mu_{ij}^{m}) = x_{ij}\alpha^{m}$$
$$ii)\ g(\mu_{ij}^{c}) = \Delta_{ij} + z_{ij}a_{i}$$
$$iii)\ a_{i} \sim F_{a}(0, D)$$
$$iv)\ Y_{ij}^{c} = (Y_{ij} \mid a_{i}) \sim F_{Y}c(\mu_{ij}^{c}, u)$$

i) Mean model

ii) Association model

iii) Latent Effects Distribution

iv) Conditional Response Distribution

Where,

Y_{ij} is the j^{th} cluster in the i^{th} pregnant women (j = 1--------- n_i ; i = 1…. N).

g is a link function for the marginal and conditional means

μ_{ij}^{m} = E (Y_{ij}) and μ_{ij}^{c} = E ($Y_{ij}|a_i$), effects of the explanatory variables.

x_{ij} are modeled through the px1 vector of marginal parameters α^m.

The vector a_i is a qx1 set of subject-specific latent effects with qxq covariance matrix D and distribution F_a (.), the function Δ_{ij} connects the marginal and conditional models as described below, and the conditional observations independently follow an exponential family distribution with mean and dispersion parameters μ_{ij}^{c} and υ (Michael *et al*., 2004). Every conditional model implies a marginal model via integration over the dependence structure, μ_{ij}^{m} = E (Y_{ij}) = E_a {E ($Y_{ij}|a_i$)} = E_a { μ_{ij}^{c} } and thus, Δ_{ij} forms a mapping between the conditional and marginal models as the solution to the integral equation h ($x_{ij}\alpha^m$) = $\int_a (\Delta_{ij} + z_{ij}a_i)dF(a)$, where h is the inverse link function h (.) = g^{-1}(.). Note that Δ_{ij} is dependent on the covariates, marginal parameters, and random effect specification, Δ_{ij} =Δ_{ij} (x_{ij}, α^m ,z_{ij}, F_a, D), but this notation is suppressed to simplify the exposition. To expand the model above, we formally relax the usual assumption that the marginal and conditional link functions are the same and allow possibly nonlinear effects to enter any of the marginal fixed, conditional fixed, or conditional random aspects. The marginalized multilevel model may then be formulated, (dropping subscripts and covariate dependence for brevity) as:

$$
\begin{aligned}
&i)\mu^m = h_m(\theta^m)\\
&ii)\mu^c = h_c(\theta^m, \Psi, a)\\
&iii)a_i \sim F_a(\Psi)\\
&iv)Y^c \sim F_Y c(\mu^c, u)
\end{aligned}
$$

i) Mean model

ii) Association model

iii) Latent Effects Distribution

iv) Conditional Response Distribution

Where h_m (.) and h_c (.) are possibly distinct inverse-link functions for the marginal and conditional means, θ^m are marginal parameters of interest, and the random effects are assumed to follow a distribution indexed by parameters Ψ . Often the latent effects and conditional response distributions are implicitly stated and the MMM may be specified with i) & ii) alone. The marginal and conditional models are tied together via integration over the random effects distribution, thus inducing the marginalization constraint

$$\begin{aligned}\theta^m &= h_m^{-1}(\mu^m) \\ &= h_m^{-1}(\int \mu^c dF_a) \\ &= h_m^{-1}\{\int h_c(\theta^m, \Psi, a) dF_a\}\end{aligned}$$

3.3.5.1. A Logistic-Probit-Normal Model

Instead of the logistic-normal conditional model for binary data, consider a Probit-normal model, as commonly used in the econometrics literature:

Probit-Normal GLMM:

$$\begin{aligned}&i)\ \Phi^{-1}(\pi_{ij}^c) = x_{ij}\alpha^c + z_{ij}a_i \\ &ii)\ a_i \sim MVN(0, D) \\ &iii)\ Y_{ij} \mid a_i \sim Binomial(n_{ij}, \pi_{ij}^c)\end{aligned}$$

$\Phi(.)$ is the cumulative normal distribution function. A marginalized version of this model may be written with definition (1) as:

Probit-Probit-Normal (PPN) MMM:

$$\begin{aligned}&i)\pi_{ij}^m = \Phi(x_{ij}\alpha^m) \\ &ii)\pi_{ij}^c = \Phi(\Delta_{ij} + z_{ij}a_i) \\ &iii)a_i \sim MVN(0, D)\end{aligned}$$

To estimate α^m , we again determine the $\Delta_{ij\ connecting}$ the mean and association models using the marginalization constraint:

$$\alpha^{m} = (x_{ij}^{'} x_{ij})^{-1} x_{ij}^{'} \Phi^{-1} \left\{ \int_{a} \Phi(\Delta_{ij} + z_{ij} a) dF_{a} \right\}$$

$$= (x_{ij}^{'} x_{ij})^{-1} x_{ij}^{'} \Phi^{-1} \left\{ \Phi \left(\frac{\Delta_{ij}}{\sqrt{1 + z_{ij}^{'} D z_{ij}}} \right) \right\}$$

and thus, $\Delta_{ij} = (\sqrt{1+z_{ij}^{'} D z_{ij}})\, x_{ij}\alpha^{m}$ In the special case where $z_{ij}a_i = a_i$ (a scalar 'random intercept' model), the conditional predictor is a simple rescaling of the marginal predictor, $\Delta_{ij} = (\sqrt{1+\tau^2}) x_{ij}\alpha^{m} = x_{ij}\alpha^{c}$ but this does not hold for general $z_{ij}a_i$. A considerable advantage of using the Probit link is that the Probit-normal marginalization integral has a closed form solution, while the logit-normal integral does not. Suppose we prefer to use a logistic regression structure for the marginal mean model but wish to retain the computational advantages of the Probit-normal association model (Michael *et al*., 2004). By relaxing the common assumption that the mean and dependence parameters are on a common scale, and obtain

Logistic Probit MMM

$$i)\pi_{ij}^{m} = \exp(x_{ij}\alpha^{m})$$
$$ii)\pi_{ij}^{c} = \Phi(\Delta_{ij} + z_{ij}a_i)$$
$$iii)a_i \sim MVN(0, D)$$

Determining Δ_{ij} with the marginalization constraint

$$\alpha^{m} = (x_{ij}^{'} x_{ij})^{-1} x_{ij}^{'} \operatorname{logit}\left\{ \int_{a} \Phi(\Delta_{ij} + z_{ij}a_i) dF_{a} \right\}$$

$$= (x_{ij}^{'} x_{ij})^{-1} x_{ij}^{'} \operatorname{logit}\left\{ \Phi\left(\frac{\Delta_{ij}}{\sqrt{1+z_{ij}^{'} D z_{ij}}} \right) \right\}$$

$\Delta_{ij} = (\sqrt{1+z_{ij}^{'} D z_{ij}})\, \Phi^{-1}\{\exp(x_{ij}\alpha^{m})\}$ (Michael *et al*.,2004)

3.3.5.2. Model Selection for MMM

Model selection is one of the most frequently encountered problems in data analysis. In most observational epidemiological studies, investigators frequently attempt to construct the most desirable statistical model using the popular methods of forward, backward, and stepwise

regression (Hosmer and Lemeshow,1989). Of course knowledge of the subject matter plays an important role in model selection, but if based strictly on the data, model selection is often carried out using one of the automated procedures built into the software, of which the most popular method is perhaps stepwise model selection. These methods pose the problem of the arbitrary selection of the significance levels in allowing a variable to enter into or to be dropped from the model during the selection process (Bozdogan ,1987). There is also the problem of multiple testing that comes with fitting and refitting the model. The issue is made more complicated in the case of repeated or longitudinal data where selecting the best model means not only to select the best mean structure but also the most optimal variance covariance structure for model selection criteria, like AIC, BIC and likelihood ratio test were used (Wolfinger,1996) for GLMM and MMM.

The principle behind AIC, first developed by Kullback-Liebler information (Kullback, 1978) which is considered to be a measure of the distance between two density functions. The variance covariance structure with the smallest AIC was selected as an appropriate model.

4. Results and Discussion

4.1. Summary of Descriptive Statistics

As it has been shown in Table 4.1, the basic descriptive information revealed, summarizes the associations between the determinant factors and the anemic status among pregnant women. The total of 1277 pregnant women from nine regional states and two city administrations in Ethiopia were eligible for this study. Among these eligible mothers, 1029 (80.6%) pregnant women were anemic where as 248 (19.4%) were non anemic.

Table 4.1 Descriptive statistics for anemic status of pregnant women associated with its related factors

Region	Non anemic in (%)	Anemic (%)	Total
Tigray	34(24.6)	104(75.4)	138
Afar	31(29.5)	74(70.5)	105
Amhara	40(19.2)	168(80.8)	208
Oromiya	23(16.2)	119(83.8)	142
Somali	50(22.6)	80(77.4)	130
B.Gumz	25(23.4)	82(76.6)	107
SNNP	36(21.6)	131(78.4)	167
Gambela	17(21.8)	61(78.2)	78
Harar	35(21.8)	63(78.2)	98
AddisAbab	6(4.3)	134(95.7)	140
Dire Dawa	15(11.9)	50(88.1)	65
Total	248(19.4)	1029(80.6)	1277(100%)

Table 4.1. Continued

Effect/variables	Category	Non anemic in (%)	Anemic in (%)	Total
Age	15-26	98(20.5)	232(79.5)	386
	27-38	210(21.1)	345 (78.9)	703
	39-50	55(29.5)	89 (47.1)	188
Smoking status	Non smoked	195 (21.8)	699(78.2)	894
	Smoked	53(13.8)	330(86.2)	383
Religion	Orthodox	105 (19.3)	439(80.7)	544
	Catholic	4(33)	8(66.7)	12
	Protestant	42(21.5)	153((78.5)	195
	Muslim	88(17.4)	419(82.6)	507
	traditional	9(47.4)	10(52.6)	19
Wealth	Poorest	77(25.3)	227(74.7)	304
	Poor	21(12.3)	150(87.7)	171
	Middle	29(21.1)	132(82.0)	161
	Rich	29 (15.6)	157(84.4)	186
	Richest	92(20.7)	363(79.8)	455
Marital status	Single	11(7.6)	134(924)	145
	Married	188(21.6)	684(78.4)	872
	Long lwp	18(18.4)	80(81.6)	98
	wioowed	13(19.4)	54 (80.6)	67
	Divorced	13(20.3)	51(79.7)	64
	No longer lwp	5(16.1)	26(83.9)	31
Edu.status	No education	78(11.4)	604(88.6)	682
	Primary	122(285)	306(71.51)	428
	Secondary	19(27.1)	51 (72.9)	70
	Higher	29(29.9)	68(70.1)	97
Occupation	No employed	101(19.8)	410(80.2)	511
	Employed	248(19.4)	1029(80.6)	766
HIV Status	HIV-	200(11.6)	1056(88.4)	1256
	HIV+	6 (12.5)	14(87.5)	20
Residence	Urban	100(20.9)	379(79.1)	479
	Rural	88(11.8)	660(88.2)	797
	Non vi intake	94(18.3)	420(81.7)	514
	Vit intake	248 (18.6)	649(81.4)	763

As it has been shown in Table 4.1, the basic descriptive statistics presents the information that summarizes the associations between the determinant factors and anemic status of pregnant women.

The percentage of anemia of pregnant women is relatively larger which is 79.5% for age groups (15-26) as compared to the youth(27-38) and relatively older age group(49-50) pregnant women are respectively,78.9%and 47.1%.

Similarly, the anemic status of pregnant women is varied with place of residence, as it can be seen in the above table 4.1, high proportion of anemic pregnant women that is a remarkable

variation of anemic due to place of residence of pregnant women. Here, the high percentage anemic status of pregnant women in rural is 81.4% and 79.1% is urban.

The percentage of anemic is 88.6% for non educated mothers, 71.5% for primary educated mother and 72.9% for mothers whose education level is secondary and (70.1%) for higher.

The anemic status of pregnant women also varied with religion, that is the percentage of anemic pregnant women for Orthodox religion are 80.7%, 33.3% for catholic 78.5% for Protestant mothers, 82.6% for Muslim and 52 % traditional. The anemic status of pregnant women is also associated wealth index. Pregnant Mother's with the lowest and highest wealth index are 74.7% and 87.7 % anemic respectively

The anemic status of pregnant women also varied with occupation, HIV statues and vitamin consumption, that is the percentage of anemic pregnant women who employed are 80.6%and80.2% for non employed, the percentage of anemic status pregnant women with HIV- is79.8% and 81.7%. for (HIV+),the percentage of anemic women whose consumed vitamin is71.8% and 80.5% for those who did not vitamin intake.

The anemic status of pregnant women also varied with smoking status and marital status, that is the percentage of anemic pregnant women who smoked is 86.2 % and 78.2% for non smokers, the percentage of anemic status for pregnant women who were single is 92.4% , 78.4%. for married, 81.6%for living together, 80.6% for widowed , 79.7% for divorced and 83.9% for no longer living with partner .

High percentage of anemic status for pregnant women has recorded in Addis Ababa, Dreidwa, Oromiya and Amhara respectively and the low percentage of anemic status recorded in Afar region. Totally (in all region), 80.6 % of pregnant women are anemic where are 19.4% are non anemic.

4.2. Statistical Analysis

4.2.1. Analysis of Data using GLM

The only difference in terms parameter estimation from generalized estimate equation (GEE), GLM and generalized linear mixed model (GLMM) is that correlation and random effect does not take in to account in GLM, however the regression coefficient interpretation is similar to generalized estimate equation since both of them are interpreted in terms of odd ratio and the output was given in Table 4.2

Effect	category	Estimate (sd.error)	95%conf Limits	OR	Pr >ChiSq
Intercept		0.8970(0.5837)	(-0.1351 2.4411)	2.452235	0. 0549
	15-26	-0.1863(0.5491)	(0.5545 0.6915)	0.8300246	0.0369
Age	27-38	0.0046(0.1896)	(0.3670762 0.37062)	1.004611	0.015
	39-50 (ref)	--------	----------	--------	-----
	orthodox	0.2451(0.1567)	(-0.7132 1.2463)	1.277749	0.1365
	catholic	-0.2243(0.9530)	(-1.1536 1.6781)	0.7990754	0.7950
Rel	Protestant	0.3828(0.4340)	(-0.5664 1.4175)	1.466385	0.1312
	Muslim	1.1756(0.5759)	(0.2613 2.1362)	3.240086	0.0001
	traditional (ref)	----------	--------- ------	----------	------
	urban	-0.1513(0.1429)	(0.1310 0.6913)	0.8595898	<.0001
Resid	Rural(ref)	---------	-----	---	------
	HIV:H+	0.1396(0.1165)	(0.0712 0.1402)	1.149814	<.0001
HIVstatus	HIV-(ref)	----------	----------	---	----
	Non Smoked	-1.0693(0.2196)	(-1.5171 -0.5817)	0.3432487	<.0001
Smoking	smoked(ref)	----------	----------	---	-----
	Non Employed	-0.0293(0.1436)	(0.2926 0.3252)	0.9711251	0.0134
Occup	Employed (ref)	----------	--------	----	-------
	Single	-0.2843(0.3214)	(-0.1134 0.6312)	0.7525408	0.4323
	Married	-0.8646(0.2504)	(-1.5945 -0.4346)	0.42122	<.0001
Mar. Status	Long lwp	-0.2564(0.1184)	(-1.1532 0.7153)	0.7738324	0.1625
	Widowed	-0.1213(0.5260)	(-1.1243 0.8123)	0.8857682	0.1915
	Divorced	0.2572(0.5344)	(-0.8471 1.5632)	1.293304	0.5503
	No longer lwp (ref)	----------	----------	------	------
	No education	2.1261(0.5419)	(1.2365 3.3652)	8.382113	<0001
Edu. status	Primary	0.1346(0.4317)	(0.1259 1.4482)	1.144079	0.0127
	Secondary	0.3411(0.6417)	(-0.3846 1.1686)	1.406494	0.0500
	Higher (ref)	----------	----------	----------	-------
	Poorest	-0.1864(0.1231)	(-0.4563 0.1923)	0.8299416	0.6034
	Poor	0.5134(0.4067)	(0.1923 1.5521)	1.670963	0.0132
Winx	Middle	0.3132(0.3151)	(-0.19621 0.3313)	1.367795	0.3612
	Rich	0.2373 (0.1912)	(-0.1834 0.6352)	1.267821	0. 5034
	Richest(ref)	----------	----------	--------	-------
Log Likelihood=-480.5		AIC=1257. 41		BIC=1378.23	

Table 4.2. Parameter estimates and Standard error under GLM

As it has been shown in Table 4.2 stands for parameter estimates and their corresponding standard errors beside the p-values for GLM model. Each parameter β_j reflects the effect of factor X_j on the log odds of the probability of pregnant women being anemic, statistically controlling all the other covariates in the model. Then, the odds ratio of variables is calculated as the exponent of β_j like GEE i.e. odds ratio = exp (β_j) disregarding correlation structure.

On Table 4.2 shown that, the religion under the category for Muslim .smoking status, educational status (all), wealth index (poor) and marital status (married) are significant effect relative to its reference since the corresponding p-value is less than 5% level of significance. And the model parameter would be interpreted using odd ratio, for instance the odd of pregnant women being anemic and whose religion Muslim given whose religion is traditional is exp(1.1756)=3.240086, means that the probability that the pregnant women being anemic and whose religion was Muslim is 24% times less likely than those pregnant women whose religion was traditional. Similarly, the odd ratio of the pregnant women being anemic and whose economic status poor given those whose economic status was richest had exp (0.5134) = 1.670963 (95%CI: 0.1923 1.5521), which means that the probability that the pregnant women being anemic and whose economic status is poor is 67 %times more likely to exposed to anemic than those whose economic status is richest keeping other covariates constant in the model. Similar interpretation would have for the remaining covariates.

Table 4.3 parameter estimates and its standard error for Beta binomial and GLM

Effects		Beta-binomial(BB) Estimate (Std. Error)	GLM Estimate (std.error)
Intercept		-1.386573(0. 20749)	-1.7203(0.3991)
HIV stataus:HIV+		-0.574046(0.36973)	-0.2243(0.9530)
Marital status : married		-0.870137(0.36788)	-0.8646(0.2504)
Religion: Muslim		0.764215(0.69911)	1.1756(0.5759)
Wealth index: poor		0.432157(0.38081)	0.5134(0.4067)
Edu.status : no education		-1.454894(0.20341)	2.1261(0.5419)
Occup : non employed		-2.116074(0.41977)	1.5740(0.3697)
Residence: urban		-0.2564(0.5184)	2.1161(0.4198)
Smoking status: smoked		1.078437(0.29564)	-1.0693(0.2196)
rho	ρ	0.1999554	----------
		QIC=1321.31	Log Likelihood=-480.5

Here, we have the two linear predictors in beta binomial model which are logit (mu), and logit (rho). The Intercept for BB value is the logit of rho. Thus, rho is estimated as the inverse logit of intercept. Therefore, rho would be estimated using the function logit inverse of (intercept value BB=0.2456), which indicates that there is over dispersion since rho greater than zero. From the above table 4.3, the standard error for GLM was slightly less than BB; this may be due to over dispersed data. Because they have different method of parameter estimation for GLM and BB. For instance, the parameter estimation for BB is quasi likelihood where as maximum likelihood for GLM. Therefore, we can see the two models in case of variation.

4.2.2. Analysis of data using (GEE)

With this regarded data (anemic status of women), marginal models (generalized estimating equation) was used to analyze the data to handle the correlated cluster data. The categorized anemic status (Hb<11 g/dL is classified under anemic and those with if Hb>=11 g/dL is considered to be non anemic) data, based on World Health Organization (WHO) cut off point, has been analyzed using the generalized estimating equation. With this analysis, GEE has considered different correlation structures such as independence and exchangeable correlation structures and compared with their QIC values. Generalized estimating equations, the user may convey a correlation structure that is often called a working correlation matrix. Before selecting the correct correlation structure, consider the model building strategy (variable selection).

The full logit model for anemic status for pregnant women of i^{th} pregnant mother from j^{th} cluster (π_{ij}) has been fitted as

$$logit(\pi_{ij}) = \beta_0 + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 edu.st_{Pr} + \beta_5 edu.st_{noedu+} + \beta_6 Religion_{or}$$

$$+ \beta_7 Religion_{pr} + \beta_8 Religion_{ca} + \beta_9 Religion_{tr} + \beta_{10} Mar.status_{single} + \beta_{11} Mar.status_{married}$$
$$+ \beta_{12} mar.status_{llwp} + \beta_{13} Mar.status_{widowed+} \beta_{14} Mar.status_{divorced+}$$
$$+ \beta_{15} Smoking_{smoked} + \beta_{16} Occupation_{employed} + \beta_{17} HIV.status_{HIV+}$$
$$+ \beta_{18} Vitamin_{intake} + \beta_{19} Wealth_{poorest} + \beta_{20} Wealth_{cpoor} + \beta_{21} Wealth_{middle}$$
$$+ \beta_{22} Wealth_{rich} + \beta_{23} Residence_{urban}$$

After fitting the model, covariates with the largest p-value of Wald test is removed and refitted the model with the rest of the covariates sequentially. Then, vitamin consumption region and some interaction covariates were excluded from the model and the remaining covariates were included in the model. Independent and exchangeable correlation structures were considered and compared to select best correlation structure depending on the QIC value.

Table 4.4: Different correlation structures with its QIC for GEE

Correlation structure	QIC value
Independent	1171.735
Exchangeable	1166.0669

As it can be seen from table 4.4, the QIC value (1166.6694) of the model with exchangeable was less than independent correlation structure and it has been selected for fitting the model as compared with independent. Thus the exchangeable correlation structure was regarded as better to fit the given model. Then now let's compare the empirical and model based standard error of independent correlation structure to fit the appropriate model:

As it can be shown in APPENDX II the standard error of the Empirical -Based Standard Error Estimates is relatively less as compared to Model-Based Standard Error Estimate. Therefore, the parameter estimates and their corresponding empirically corrected standard errors with the p-values from the final GEE model for parameter estimate was parsimonious and given in Table below.

Table 4.5 Parameter estimates (empirically corrected standard errors) for GEE

Effect	category	Para	Estimate (sd.error)	95%conf Limits	OR	P r > |Z|
Intercept		β_o	0.9870(0.5837)	(-0.1571 2.1311)	2.683	0. 0909
Age	15-26	β_1	-0.1936(0.2490)	(0.2945 0.6817)	0.824	0.0369
	27-38	β_2	0.0046(0.1896)	(0.3670762 0.37062)	1.005	0.005
	39-50 (ref)	-	--------	----------	----	-----
Rel	orthodox	β_3	0.4202(0.4186)	(-0.4002 1.2406)	1.522	0.3155
	catholic	β_4	-0.2240(0.8210)	(-1.8332 1.3852)	0.799	0.7850
	Protestant	β_5	0.3928(0.5330)	(-0.6519 1.4375)	1.481	0.4612
	Muslim	β_6	1.1844(0.4759)	(0.2517 2.1172)	3.269	0.0128
	traditional (ref)	---	----------	----------	------	------
Resid	urban	β_7	-0.1485(0.1528)	(0.1510 0.4479)	0.862	0.0012
	Rural(ref)	--	---------	-----	---	----
HIVstatus	HIV:H	β_8	0.1450(0.1047)	(0.0601 0.3502)	1.39	<.0001
	HIV-(ref)	---	----------	----------	---	----
Smoking	Non Smoked	β_9	-1.0784(0.2595)	(-1.5871 -0.5697)	0.340	<.0001
	smoked(ref)	-	----------	----------	---	-----
Occup	Non Employed	β_{10}	-0.0367(0.1336)	(0.2986 0.3252)	0.964	0.0034
	Employed (ref)	---	----------	--------	---	-------
Mar. Status	Single	β_{11}	-0.2944(0.4288)	(-1.1348 0.5460)	0.745	0.4923
	Married	β_{12}	-0.9646(0.2704)	(-1.4945 -0.4346)	0.381	0.0004
	Long lwp	β_{13}	-0.2664(0.4484)	(-1.1453 0.6125)	0.766	0.5525
	Widowed	β_{14}	-0.1111(0.5160)	(-1.1224 0.9002)	0.895	0.8295
	Divorced	β_{15}	0.2672(0.6081)	(-0.9247 1.4591)	1.306	0.6604
	No longer lwp (ref)	---	----------	----------	--	----
Edu. status	No education	β_{16}	2.1161(0.4617)	(1.2112 3.0210)	8.2989	<0001
	Primary	β_{17}	0.6612(0.4016)	(-0.1259 1.4482)	1.937	0.0997
	Secondary	β_{18}	0.5420(0.3197)	(-0.0846 1.1686)	0.821	0.0900
	Higher (ref)	--	----------	----------	---	----
Winx	Poorest	β_{19}	-0.1963(0.1931)	(-0.5749 0.1822)	0.821	0.3094
	Poor	β_{20}	0.7358(0.3062)	(0.1357 1.3359)	2.087	0.0162
	Middle	β_{21}	0.2123(0.2036)	(-0.1867 0.6114)	1.237	0.2970
	Rich	β_{22}	0.2768(0.2175)	(-0.1496 0.7032)	1.319	0.2033
	Richest(ref)	---	----------	----------	----	-----
Corr.			0.0253504575			

The parameter estimates for GEE stand for the effect of the predictors averaged across all individuals with the same predictor values. Like standard normal logistic regression, the interpretation of the parameters in the marginal (population average) model would be interpreted in terms of odd ratio. The final proposed reduced model for GEE is:

$$logit(\pi_{ij}) = \beta_0 + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 Religion_{musilum} + \beta_4 Residence_{urban} + \beta_5 smokinhg\ status_{non\ smoked} + \beta_6 Marital\ status_{married} + \beta_7 Edu.statusn_{no\ education} + \beta_8 Wealth\ poor + \beta_9 Occupation_{employed} + \beta_{10} HIV_{HIV+}$$

As it has been seen in Table 4.5, it stands for the parameter estimates and their corresponding empirically corrected standard errors beside the p-values for GEE model. Each parameter β_j reflects the effect of factor X_j on the log odds of the probability of pregnant women being anemic, statistically controlling all the other covariates in the model. Then, the odds ratio of variables were calculated as the exponent of β_j i.e. odds ratio = exp (β_j).

The GEE analysis from table 4.5 shows that, age is significantly related to anemic status of pregnant women .The odds ratio of anemic pregnant women whose age 15 to 20 had exp(β_1)= exp(-0.1936)= 0.824(95% CI: 0.6817 0.2945) times lower than those pregnant women whose age group (40-49), which means that the probability that the pregnant women being anemic whose age 15to 20 is 17.6% times less likely to be anemic than those anemic pregnant women whose age group (40-49) in the same j^{th} cluster.

Similarly, some part of religion is statistically significant on anemic status of pregnant women since not all religion are not statistically significant Thus, .the odds ratio of anemic pregnant women whose Religion is Muslim had exp (1.1844)=3.268725 (95% CI: 0.2517 2.1172) times higher than pregnant women whose religion is traditional. Likewise, residence is one of factors that related to anemic status of women , which means that the pregnant women who lived in urban had exp(-0.1485)= 0.862 (95%CI:0.4479 0.1510)times lower than those pregnant women being anemic who lived in rural ,which means that the probability that the pregnant women who lived in urban and being anemic is 13.8% less likely than those who lived in rural and being anemic**.**

Smoking status is also related to anemic status of pregnant women. The odds ratio of anemic pregnant women who did not smoked is exp (-1.0784)=0.3401393 (95%CI:-1.5871 -0.5697) times lower than pregnant women being anemic who did smoke. Equivalently, the probability of anemic women who has not smoked is 65.98% less likely than those who smoked cigarette.

Marital status also has significantly associated with anemic status of pregnant women. The odds ratio of pregnant women being anemic who were married is exp(-0.9646= 0.3811356(95%CI:-1.4945 -0.4346)times lower than those pregnant women being anemic who did not live with partner .Equivalently, the probability of pregnant women who are married and being anemic is 61.8% times less likely than those women who did not live with partner.

In similar fashion, wealth index and education status have an effect for the anemic status of pregnant women, the odd ratio of the pregnant women being anemic and whose income is poor

is exp(0.7358)= 2.087151(95%CI:0.1357 1.3359) times higher than pregnant women who is anemic and who are richest. Similarly the odd ratio of the pregnant women who is being anemic and there education level is illiterate is exp (2.1161) = 2.1161 (95%CI:1.2112 3.0210) times higher than those women who have higher education level.

The odds of being anemic pregnant women who are HIV+ is exp(0.1450)= 1.15604 (95%CI:0.0601 0.3502)times higher than pregnant women who are anemic and HIV-, equivalently, those women who is anemic and HIV+ is 15.6 % more likely at risk than the reference group. Likewise, the odd ratio of the anemic status of pregnant women who were smoked is exp(-1.0784)= 0.3401393 (95%CI:-1.5871 -0.5697) times lower than pregnant women those who were not smoked or the probability that the women who are smoked and being anemic is 65.9% times less likely than the reference group(those who were not smoked).Working status also has significant effect on anemic status of pregnant women, that is the odd ratio that the women being anemic and who were employed is exp(-0.0367) = 0.9639653 (95%CI:0.2986 0.3252) times lower than those who were not employed or the probability that the pregnant women being anemic and employed is 3.6% less likely anemic than those who were not employed.

4.2.3. Analysis of Generalized Linear Mixed Model (GLMM)

4.2.3.1. Model Building in GLMM

Generalized Linear Mixed Models is mainly extension of generalized linear models to correlated data, generalized mixed models to discrete outcome data and likelihood estimation is computationally challenging. Furthermore, the model also included the random effects in this case, random intercepts to address the between and within-regional variations. Then the saturated models with the two random intercepts associated with covariates were fitted. GLMM were fitted as follows where, b_j & b_{ij} two random intercepts.

$$
\begin{aligned}
logit(\pi_{ij}) = {} & \beta_0 + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 edu.st_{Pr} + \beta_4 edu.st_{Secd+} \\
& + \beta_5 edu.st_{noedu+} + \beta_6 Religion_{or} + \beta_7 Religion_{pr} + \beta_8 Religion_{ca} + \beta_9 Religion_{tr} \\
& + \beta_{10} Mar.status_{single} + \beta_{11} Mar.status_{married} + \beta_{12} mar.status_{llwp} \\
& + \beta_{13} Mar.status_{widowed+} \beta_{14} Mar.status_{divorced+} + \beta_{15} Smoking_{non\ smoked} \\
& + \beta_{16} Occupation_{non\ employed} + \beta_{17} HIV.status_{HIV+} \\
& + \beta_{18} Vitamin.consum_{vit\ intake} + \beta_{19} Wealth_{poorest} + \beta_{20} Wealth_{cpoor} \\
& + \beta_{21} Wealth_{middle} + \beta_{22} Wealth_{rich} + \beta_{23} Residence_{urban+} bj + bij
\end{aligned}
$$

In order to decide on the better of the two random effects models, two models were fitted, one the saturated model above with two random intercepts to estimate between and within regional variations and the other with one random intercept model to estimate within regional variation. AIC and Likelihood ratio test (LRT) were used to compared the two models to select an appropriate models.

4.2.3.2. Parameter Interpretation of GLMM

In the GLMM data analysis, the parameter interpretation is based on subject specific or cluster effect as well as fixed effect. The parameter interpretation is conditional on the random effects, which is common for all individual pregnant women in the same cluster.

Table 4.6 Parameter estimates (standard errors) and corresponding P value for GLMM.

Parameter	Estimate (s.error)	t Value	Pr > |t|	Alpha	95%conf.limt	
beta0	1.2393(0.5206)	2.38	0.0208	0.05	(0.1959	2.2827)
beta1	-0.1356(0.07058)	-1.92	0.0098	0.05	(0.00580	0.2771)
beta2	0.05336(0.07058)	0.76	0.0382	0.05	(0.08808	0.1948)
beta3	0.01947(0.08256)	0.24	0.0044	0.05	(0.1460	0.1849)
beta4	0.08947(0.1720)	0.52	0.0051	0.05	(0.2553	0.4342)
beta5	-0.1005(0.1691)	0.59	0.0048	0.05	(0.2384	0.4394)
beta6	1.4567(0.2274)	6.41	<.0001	0.05	(1.0010	1.9125)
beta7	-0.2130(0.1359)	-1.57	0.1227	0.05	(-0.4853	0.05928)
beta8	2.1161(0.1567)	2.56	<.0001	0.05	(-1.7563	-0.1849)
beta9	0.05447(0.05239)	1. 04	0.3030	0.05	(-0.05052	0.1595)
beta10	0.08020(0.1574)	0.51	<.0000	0.05	(0.2353	0.3957)
Sigma	1.0668(0.1653)	6.46	<.0001	0.05	(0.7357	1.3980)
-2 Log Likelihood =1148.3		AIC =1172.3			BIC=1196.6	

Where beta of 0,1,2,3,4,5,6,7,8,9 and 10 represent the parameter estimate for interecept,age category 15to20;21 to 39,muslium,urban,HIV+,non smoked, married, Non educated , poor, and non employed respectively.

Under GLMM model, the parameter of random effect is not estimable, but we can estimate in terms of variability. Thus, the estimates of standard deviation of random effect is 1.0668 associated with small p-value ($p<0.0001$), and which indicates that there is significance heterogeneity within and between regions, since it differ from zero, on the anemic status of pregnant women.

Parameter of GLMM would be interpreted in the following: to illustrate the difference in interpretation, consider the effect of age on the probability of being anemic using the generalized linear mixed model had exp (-0.1356) = 0.8731918 time of the anemia testing lower to develop anemia than if those individuals who aged 40 to 49. Similarly, the odd of being anemic among

pregnant women whose religion was Muslim had exp (0.01947) = 1.019661times higher than the individuals those whose religion was traditional or the probability that the pregnant women being anemic whose religion was Muslim was 1.9% times more likely exposed to anemic than those whose religion was traditional within in the same cluster . The interpretation of other predictor variables can be done in a similar way**.**

4.2.3.3. Mode1Diagnostic for GLMM

As it has been shown in APPENDXI, the Q-Q plot from the following figure in first panel verifies that the residuals are close to normally distributed and symmetric around zero. Thus, it meets the assumption of the distribution of error terms. As well, to the above, the non linearity of the Q-Q plot confirms the model is not linear. Residuals versus observation CLID number plot panel two, also suggested that the residuals are symmetric around zero (i.e. positive and negative residuals are almost equal). Q-Q plots for normality of random effects at regional and cluster levels are also given in the figure and illustrates that the random effects are normally distributed with mean zero and variance covariance matrix D. Thus, the fitted GLMM model is well for the given data.

4.2.4. Analysis of Data using MMM

The parameter of the marginalized multilevel model, Logit-Probit-Normal MMM for binary data was used due to its closed form solution.

Table 4.7: Parameter estimates of marginalized multilevel model

Parameter	Estimate (std.error)	t Value	Pr > |t|	Alpha	95% conf.int limit	
beta0	1.1393(0.4792)	2.38	0.0209	0.05	(0.1789	2.0997)
beta1	0.1389(0.06520)	2.13	0.0377	0.05	(-0.2695	-0.0082)
beta2	0.05013(0.06520)	0.77	0.0452	0.05	(0.08053	0.1808)
beta3	0.01406(0.07660)	1.18	0.0025	0.05	(0.1394	0.1676)
beta4	0.08332(0.1609)	2.52	0.0066	0.05	(0.2391	0.4058)
beta5	-0.1020(0.1575)	-0.65	0.5200	0.05	(-0.4177	0.2137)
beta6	1.3566(0.2109)	6.43	<.0001	0.05	(0.9340	1.7792)
beta7	-0.1891(0.1221)	-1.55	0.0 271	0.05	(0.0555	0.4338)
beta8	2.1161(0.5321)	1.65	<.0001	0.05	(-0.3453	-0.1235)
beta9	0.05030(0.04841)	1.04	0.0033	0.05	(0.04671	0.1473)
beta10	0.07259(0.1460)	0.50	0.6211	0.05	(-0.2200	0.3652)
tau	0.3714(0.05361)	6.93	<.0001	0.05	(0.2640	0.4788)

-2 Log Likelihood =1150.2 AIC=1174.2 BIC =1198.5

Where beta of 0,1,2,3,4,5,6,7,8,9 and 10 represent the parameter estimate for interecept,age category 15to20;21 to 39,muslium,urban,HIV+,non smoked, married, Non educated , poor, and non employed respectively.

Parameter estimation for MMM: There is a critical distinction between the marginal parameter $\beta^M$and the conditional parameterβ^C. The conditional regression coefficient β^C contrasts the expected response for different values of the measured covariates $X_{ijk}$for equivalent values of the latent variableb_{ijk}.The marginal coefficient does not attempt to control for the unobserved b_{ijk} . As the result of this, the interpretation of β^Ccan be particularly difficult for multilevel models with level-2 for covariates since no direct matching of b_{ijk} is observed for these contrasts. However, if the variance of the latent variable is independent of *X*, then the marginal and conditional model structures will be the same. With this assumption, the model parameter can be interpreted marginally.

The mean of the pregnant women who aged 15 to 20 were being anemic decreased by 0.1936 as compared to those anemic pregnant women whose age 40 to 49.Similarly, the mean of the pregnant women who aged 21 to 39 were being anemic decreased by mean of the pregnant women who aged 15 to 20 were being anemic increased by 0.1936 as compared to those anemic pregnant women whose age 40 to 49.

The mean of the pregnant women whose religion Muslim were being anemic increased the risk by 1.1844 as compared to those anemic pregnant women whose religion was traditional.

Table 4.8: Parameter estimates, standard errors for three models GEE, MMM and GLMM

		GEE	MMM	GLMM
Effects	parameter	estimate (s.e)	estimate (s.e)	estimate (s.e)
Intercept	beta0	0.9870(0.5837)	0.1393(0.4792)	1.2393(0.5206)
Age: (15-20)	beta1	-0.1936(0.2490)	-0.1389(0.06520)	-0.1356(0.07058)
Age :(21-39)	beta2	0.0046(0.1896)	0.05013(0.06522)	0.05336(0.07058)
Religion: Muslim	beta3	1.1844(0.4759)	0.01406(0.07660)	0.01947(0.08256)
Residence: urban	beta4	-0.14185(0.1528)	0.08332(0.1609)	-0.1005(0.1691)
HIV: HIV+	beta5	0.1450(0.1047)	0.1020(0.1575)	-0.1005(0.1693)
Smoking: non smoked:	beta6	-1.0784 (0.2595)	1.3566(0.2109)	1.4567(0.2274)
Marital status: married	beta7	-0.9646(0.2704)	-0.1891(0.121)	-0.2130(0.1359)
Edu.status: no education	beta8	2.1161(0.4617)	2.1161(0.5321)	0 .05447(0.05239)
Wealth index: poor	beta9	-0.0367(0.3062)	0.07259(0.1460)	0.05447(0.05239)
Work status: non employed	beta10	-0.367(0.1336)	0.053030(0.1460)	0.08020(0.1574)
Variation	d	--------	0.3714(0.05361)	1.0668(0.1653)
Corr.	ρ	0.0252504775	----------	-----

From table 4.8, the standard errors of MMM are almost small comparing with GEE but the estimated values of GEE and MMM are almost similar. Taking MMM with regard to precision as compared to GEE for estimates for population average interpretations. On the other hand, the variation within region for MMM is less than GLMM, which indicates the MMM is better to fit the data, however, due to the relatively high random-effect variance, the GLMM and GEE estimates are quite more different, with the estimates from the MMM lying in between. Therefore, MMM is the best or robust parameter estimates by combing marginal and conditional random effect in the data.

4.3. Discussion

Pregnant women with higher probability of occurrence of these determinant factors would be inferred to be most likely to experience anemia as hemoglobin (Hb) concentration below the normal level is often associated with anemia. Various studies, the data were analyzed using descriptive statistics as well as binary logistic regression .This may not give valid inference since relevant information will not take in to account. Thus, the analysis was extended to other statistical methods to account for the clustered nature of correlated observations. The data were then analyzed using the following model families: Generalized linear model (GLM), beta-binomial model, Marginal models (GEE), Generalized Linear Mixed Model (GLMM) and Marginalized Multi level Model (MMM). On the other hand, women who lived in urban had the lowest probabilities of developing anemia. Even if different family models have fitted in the same analysis, it should be kept in mind that those model families are rather different, and that the parameters have to be interpreted differently since some models did it account random effect where as some models didn't. In addition, indeed, different models have own method of estimation , it is due to the fact that , the model parameter would be interpreted separately , however some models would be compared so as to select the best robust models for the given data .Indeed ,in practical situations, the choice on which model family to use is guided by the research question. Furthermore, the given different marginal model family further indicated a strong significant association between any two pairs of responses as well as pairs of observations within the same cluster. In GLMM, it appeal make a note of that variable region did not appear in the final model, however, the significance of measures of associations and the presence of type of residence in the final model can provide information about within region variation of anemia.

Under the last best model analysis ,anemia and socio-demographic variables including residence, religion(Muslim), occupation, marital status(married), income status(poor) ,and educational status(no education),smoking status and age categorized showed a statistical significant difference with anemia among pregnant women ,this finding supported on multivariate logistic regression analysis on determinants of anemia in pregnant women at bushulo health center in southern Ethiopia(Bamlaku Tadege, 2009). Educational status have strongly related to the risk of anemia among pregnant women in Ethiopia, similar results would be obtained on the study conducted on risk factors of anemia during pregnancy among pregnant women in India showed a statistical significant association between education and anemia which is consistent with the

current study (Bechuram *et al*.,2006)and similarly (Dutta et al,1992) in a study reported that pregnant women with a low literacy level had significantly more from anemia compared to highly literate women. This finding indicates the need for strength ending of interventions related to education to women to create awareness of antenatal care, balanced diet during pregnancy and family planning. In all models, religion (Muslim) and poor income covariates have significant effect for facilitating in reducing hemoglobin level(increasing anemia), it may be due to the fact that, I predicted the element dictating individual eating habits is religious diets, which is quite strict and culturally significant and a general attitude associating good image to lack of eating could easily play a significant role in the high prevalence of nutritional anemia and one of the causative factors in high level of anemia found in this study could result from poor income. This finding supported by (Egbert, 1996) who had stated that income had been identified as an indicator of the quality and quantity of foods available to pregnant mothers.

The consumption of vitamin in take has no significant effect for anemia among pregnant women this study is supported by (Gebremedin, Enquoselassie. 2011), for binary logistic regression showed that the vitamin A supplement during pregnancy and postpartum period, respectively, didn't have a significant effect in reducing the burden of anemia.

The present study showed poor educational, nutritional and other health indicators during pregnancy in women of lower socio-economic status as compared to those with upper socioeconomic status. In the present study significant association was found between Income and Anemia. This study is supported by (Sharma *etal.* 2007), for chi-square test of association showed that socio economic status is found to be a major explanation for the women having anemia in their study comprising of various social status groups, categorized on the basis of family income, found that the most females from low income category were more iron deficient. Present study clearly shows that Unfavorable socio demographic factors are the major barriers to the efforts in place for the prevention of anemia during pregnancy.

5. Conclusion and recommendation

5.1. Conclusion

Anemia has moderate public health significance in Ethiopia. Pregnant women lived in rural areas, being from the lower economic ,educational status categories(no education) , marital status (married),religion(Muslim) , smoking status , working status and HIV status were important predisposing factors to anemia. All the three models led to the same conclusion that age (in year) , type of residence, educational status categories(no education) , marital status (married), religion(Muslim) , smoking status , working status and HIV status. More generally, Socio economic status, literacy of women is the major determinates that contribute to the problem of anemia. Education is the basic factor for change.

5.2. Recommendation

According to findings of this cross sectional study, place of residence, HIV status, smoking status, religion, income level are significant factors for anemia among pregnant women. Clearly, it follows due to strong association between anemia and socio-demographic factors and economic factors, this means that reproductive women aged (15-49) especially pregnant women brought to health facilities by giving awareness about anemia since the result of this study showed that low income pregnant women, poor education level and additional factors mentioned above were high risk factor for anemia so that Government should design strategies and policies to enhance women education to make them independent in socio-economic and cultural decision, which directly and indirectly affect women health status due to anemia.

Furthermore, in this analysis, we have studied how the risk of being anemic depends on age of pregnant women, type of residence, smoking status, working status, education status, marital status, and HIV and income level. However, it is worth noting that the probability of being anemic, that is, having hemoglobin (Hb) concentration below the normal level could be affected by other factors such as nutritional deficiencies, hookworm infections and inherited red blood cell disorders. Investigation of such factors could be recommended in future studies. However, challenges may stretch out on the side of resources made available and possibly means of collecting these factors.

References

Agresti, A., (2007). An Introduction to Categorical Data Analysis. John Wiley and Sons, Inc New York

Agresti, A., (2002). Categorical Data Analysis, Second Edition. Wiley Inter science. New York

Balemi,A and Lee. A (1999).Some proportesis of Liang-Zeger method applied to cluster binary Regression. *Aust.N.Z.J.Stat*.41,43-58.

Bamlaku Tadege.,(2009). Determinants of anaemia in pregnant women with emphasis on intestinal helminthes infection at bushulo health center in southern Ethiopia

Bechuram M., Vikal T., Ranjan G. (2006). Risk factors of anaemia during pregnancy among the Garo of Meghalaya, India. *Human Ecology Special Issue*, 14(1):27-32.

Bondevik, G .T. Abel (2001). Maternal hematological status and risk of low birth weight and preterm delivery in Nepal. *Actan Obstet gynaecol Scand, vol.80 p.402-8*

Box, G., Jenkins, GM. and Reinsel G (1994). Time Series Analysis: Forecasting and Control. Pretice-Hall international,Inc.

Bozdogan, H.(1987). Model selection and Akaike's information criteria (AIC): The general theory and its analytical extensions. Psychometrika 52, 345-370.

Cosmas, G., (2011).Socio-Demographic determinants of anemia among children aged 6-59 months in mainland Tanzania, University of Hasselt.

De Benoist, B.,et al. (2008). *Worldwide prevalence of anaemia 1993-2005: WHO Global Database on Anaemia.* Geneva: World Health Organization.

Diggle, P.J., Heagerty, P.J., Liang, K.Y. and Zeger, S.L. (2002). Analysis of Longitudinal Data. (2nd edition).Oxford: Oxford University Press.

Gebremedin S. (2004). Prevalence of anaemia and its predisposing factors in pregnant mothers attending antenatal care in Mettu Carl hospital.

Guyatt L., Snow W. (2001). The epidemiology and burden of Plasmodium falcip arum related anemia among pregnant women in sub-Saharan Africa. *Am. J. Trop.Med. Hyg.* 64 (*S1–2*), 36–44.

Dutta, P.K., Nagraj, T., Gopinath, V.P. 1992.A case control study of anemia in pregnancy. Indian Journal of Preventive and Social Medicine, 23(1):1-5.

Faraway, J., (2006).Extending the Linear Model with R, Generalized Linear, Mixed Effects,and Nonparametric Regression Models, Boca Raton London New York,

Heagerty, P. J. and Zeger, S. L . (2000). Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). Statistical Science 15, 1–26.

Hercberg S., Galan P. (1992). Nutritional anemia Baillieres *Clin Haematol* **5**:143 – 168.

Hosmer, D.W. and Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley and Sons,Inc,pp 376-381.

Idowu OA, Mafiana CF, Dapo S (2005) Anaemia in pregnancy: a survey of pregnant women in Abeokuta, Nigeria. Afr Health Sci 5: 295-299.

Johannessen, A.. (2008) Predictors of mortality in HIV-infected patients starting antiretroviral therapy in a rural hospital in Tanzania. *BMC Infectious Diseases* **8**:52.

Karaoglu.(2010).The prevalence of nutritional anemia in pregnancy in an East Anatolian province, Turkey.BMC public Health, 10:329.

Kidanto *et al*. (2009) Risk for preterm delivery and low birth weight are independently increased by severity of maternal anaemia. *South African Medical Journal* 99, 98-102.

Kleinman, J. (1973) Proportions with extraneous variance: single and independent samples. Journal of the American Statistical Association, 68, 46–54..

Kullback, S. (1978). Information Theory and Statistics, Massachusetts: Peter Smith..

Meda,,*et al* (1999). Anemia during pregnancy in Burkina Faso, West Africa, 1995-96: prevalence and associated factors. *Bulletin of the World Health Organization* 77, 916-922.

Michael, E. Griswold & Scott, L. Zeger (2004). on marginalized multilevel models and their computation, Johns Hopkins Bloomberg School of Public Health

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

Molenberghs, G., Verbeke, G., Demetrio, C., and Vieira, A. (2010) A family of generalized linear models for repeated measures with normal and conjugate random effects. Statistical Science, 25, 325–347.

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

Samson Gebremedin, Fikre Enquselassie.(2011).Correlates of anemia among women of reproductive age in Ethiopia: Evidence from Ethiopian DHS 2005. *Ethiop. J. Health Dev.* 2011;25(1):22-30.

Scholl TO, Hediger ML(1994). Anemia and iron-deficiency anemia: compilation of data on pregnancy outcome.*American Journal of Clinical Nutrition*, 59:492S–500S.

Sharma JB. Nutritional anaemia during pregnancy in non industrialised countries. In: *Progress in Obstetrics &Gynaecology*. 15: John Studd (ed) Churchill Livingstone2003;103-120.

Sifakis S., Pharmackides G. (2000). Anaemia in pregnancy Ann NY *Acad Sci,* **90**:125 – 136.

Singh B., Fong F, Arikumeron S. (1998). Anaemia in pregnancy, a cross sectional study in Singapore, *European J Clin Nutr*, **52**(1) 65- 70.

Skellam, J.G. (1948) A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials, Journal of the Royal Statistical Society, Series B, 10, 257–261.

Tadios Y. (1996). Prevalence of anaemia and its risk factors among pregnant mothers following

Antenatal in Asendabo health center.

Yuan Xing & Hong Yan (2009).Hb levels and anemia evaluation during pregnancy in the high lands of Tibet, Lhasa People´s Hosp.University College of Medicine, China.

Wolfinger R.D.(1996). Heterogeneous variance covariance structures for repeated measures. Journal of Agricultural, Biological, and Environmental Statistics. 1(2), 205-230.

WHO (2006) . *Antiretroviral Drugs for Treating Pregnant Women and Preventing HIV Infections in Infants: Towards Universal Access: Recommendations for a Public Health Approach*. World Health Organization, Geneva.

WHO (2011). *Hemoglobin concentrations for the diagnosis of anaemia and assessment of severity.* Vitamin and Mineral Nutrition Information System. Geneva, World Health Organization,(WHO/NMH/NHD/MNM/11.1)(http://www.who.int/vmnis/indicators/haemoglobin.pdf, accessed [date]).

World Health Organization. (1996). Report of the WHO informal consultation on hookworm infection and anaemia in girls and women. Geneva: WHO, CTD/SIP/ 96.1.

World Health Organization /UNICEF. (2001). Iron deficiency anaemia, assessment, prevention and control: a guide for program me managers. United Nations University

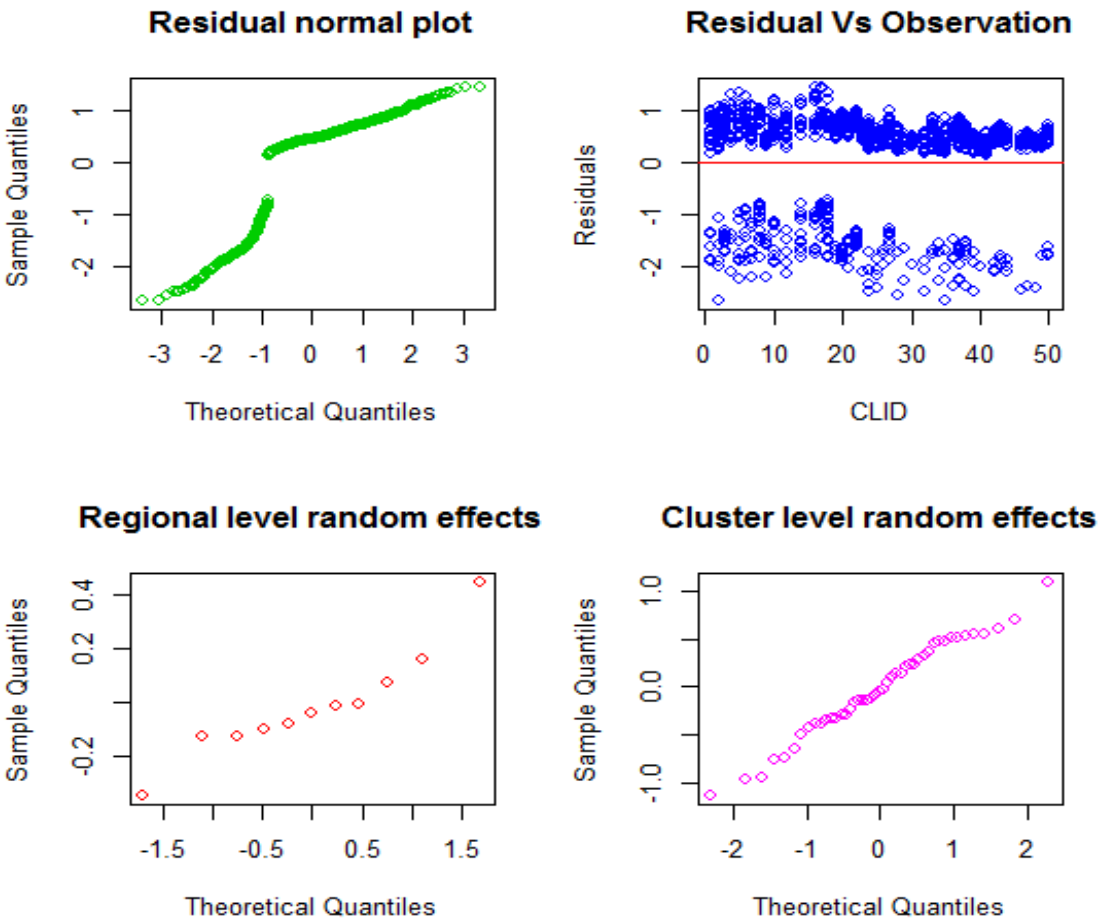
World Health Organization. (1991). Controlling iron deficiency. ACC/SCN state of the art series Nutrition Policy Discussion Paper No.9

Worldwide prevalence of anaemia, 1993–2005: WHO global database on anaemia / Edited by Bruno de Benoist.

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

APPENDEX

APPENDEX I: Normality assumption checking and diagnostic for random effect for GLMM



APPENDEXII

Exchangeable correlation structure for model based and empirical standard error for GEE

	correlation structure for GEE			
parameter	Model based	s.error	Empirical based	s.eroor
β_0	0.9870	0.8786	0.9870	0.5837
β_1	-0.1936	0.2607	-0.1936	0.2490
β_2	-0.0046	0.2117	-0.0046	0.1896
β_3	0.4202	0.5480	0.4202	0.4186
β_4	-0.2240	0.8591	-0.2240	0.8210
β_5	0.3928	0.5557	0.3928	0.5330
β_6	1.1844	0.5478	1.1844	0.4759
β_7	-0.1485	0.1574	-0.1485	0.1528
β_8	0.1450	0.1586	0.1450	0.1047
β_9	-1.0784	0.1956	-1.0784	0.2595
β_{10}	-0.0367	0.1535	-0.0367	0.1336
β_{11}	-0.2944	0.6416	-0.2944	0.4288
β_{12}	-0.9646	0.5404	-0.9646	0.2704
β_{13}	-0.2664	0.6126	-0.2664	0.4484
β_{14}	-0.1111	0.6835	-0.1111	0.5160
β_{15}	0.2672	0.6716	0.2672	0.6081
β_{16}	2.1161	0.4198	2.1161	0.4617
β_{17}	0.6612	0.4219	0.6612	0.4016
β_{18}	0.5420	0.4744	0.5420	0.3197
β_{19}	-0.1963	0.2016	-0.1963	0.1931

β_{20}	0.7358	0.2768	0.7358	0.3062
β_{21}	0.2123	0.2544	0.2123	0.2036
β_{22}	0.2768	0.2530	0.2768	0.2175

APPENDEXIII: variance covariance structure for MMM

Covariance Matrix of Parameter Estimates

Row	Parameter	beta0	beta1	beta2	beta3	beta4	beta5	beta6	beta7
1	beta0	0.2297	-0.00659	-0.00659	-0.01858	-0.04091	-0.01284	-0.01203	-0.02361
2	beta1	-0.00659	0.004251	0.004251	-0.00003	-0.00020	0.000209	-0.00006	-0.00040
3	beta2	-0.00659	0.004251	0.004251	-0.00003	-0.00020	0.000209	-0.00006	-0.00040
4	beta3	-0.01858	-0.00003	-0.00003	0.005867	-0.00040	0.000584	0.001188	0.001160
5	beta4	-0.04091	-0.00020	-0.00020	-0.00040	0.02589	-0.00070	0.000319	0.000707
6	beta5	-0.01284	0.000209	0.000209	0.000584	-0.00070	0.02482	-0.00065	-0.00063
7	beta6	-0.01203	-0.00006	-0.00006	0.001188	0.000319	-0.00065	0.04447	-0.00155
8	beta7	-0.02361	-0.00040	-0.00040	0.001160	0.000707	-0.00063	-0.00155	0.01491
9	beta8	-0.00881	0.000014	0.000014	0.000715	0.000026	-0.00060	0.000597	-0.00022
10	beta9	-0.00881	0.000014	0.000014	0.000715	0.000026	-0.00060	0.000597	-0.00022
11	beta10	-0.01215	-0.00041	-0.00041	-0.00025	0.000858	-0.00072	0.000734	0.000131
12	tau	-0.00083	0.000556	0.000556	0.000133	-0.00045	0.000237	0.002522	-0.00104

Covariance Matrix of Parameter Estimates

Row	beta8	beta9	beta10	tau
1	-0.00534	-0.00881	-0.01215	-0.00083
2	0.000023	0.000014	-0.00041	0.000556
3	-0.000018	0.000014	-0.00041	0.000556
4	0.002343	0.000715	-0.00025	0.000133
5	0.02132 7	0.000026	0.000858	-0.00045
6	0.000856	-0.00060	-0.00072	0.000237
7	-0.001358	0.000597	0.02132	0.002522
8	-0.021326	-0.00022	0.000131	-0.00104
9	-0.001848	0.000858	0.000816	0.123085
10	-0.000123	0.002343	-0.00009	0.000037

11	0.001058	-0.00009	0.02132	-9.64E-7
12	-0.001438	0.000037	-9.64E-7	0.002874

Correlation Matrix of Parameter Estimates

Row	Parameter	beta0	beta1	beta2	beta3	beta4	beta5	beta6	beta7
1	beta0	1.0000	-0.2109	-0.2109	-0.5061	-0.5306	-0.1701	-0.1191	-0.4035
2	beta1	-0.2109	1.0000	1.0000	-0.00549	-0.01905	0.02035	-0.00448	-0.05008
3	beta2	-0.2109	1.0000	1.0000	-0.00549	-0.01905	0.02035	-0.00448	-0.05008

Correlation Matrix of Parameter Estimates

Row	beta8	beta9	beta10	tau
1	-0.00549	-0.3796	-0.1736	-0.03246
2	0.02035	0.004417	-0.04313	0.1592

Correlation Matrix of Parameter Estimates

Row	Parameter	beta0	beta1	beta2	beta3	beta4	beta5	beta6	beta7
4	beta3	-0.5061	-0.00549	-0.00549	1.0000	-0.03242	0.04839	0.07356	0.1241
5	beta4	-0.5306	-0.01905	-0.01905	-0.03242	1.0000	-0.02761	0.009400	0.03601
6	beta5	-0.1701	0.02035	0.02035	0.04839	-0.02761	1.0000	-0.01967	-0.03289
7	beta6	-0.1191	-0.00448	-0.00448	0.07356	0.009400	-0.01967	1.0000	-0.06026
8	beta7	-0.4035	-0.05008	-0.05008	0.1241	0.03601	-0.03289	-0.06026	1.0000
9	beta8	-0.0276	-0.00448	-0.07365	0.01976	-0.01967	-0.02298	0.00940	-0023
10	beta9	-0.3796	0.004417	0.004417	0.1928	0.003386	-0.07836	0.05847	-0.03667
11	beta10	-0.1736	-0.04313	-0.04313	-0.02272	0.03652	-0.03116	0.02385	0.007330
12	tau	-0.03246	0.1592	0.1592	0.03249	-0.05203	0.02804	0.2231	-0.1582

Row	beta8	beta9	beta10	tau
4	-0.07836	0.1928	-0.02272	0.03249
5	0.03652	0.003386	0.03652	-0.05203
6	-0.02272	-0.07836	-0.03116	0.02804
7	0.03652	0.05847	0.02385	0.2231
8	0.02385	-0.03667	0.007330	-0.1582
9	1.0000	-0.03116	0.1928	0.2231
10	0.01423	1.0000	-0.01330	0.01423
11	0.02385	-0.01330	1.0000	-0.00012

12	-0.02298	0.01423	-0.00012	1.0000