2. Methods

2.1 Study Population

This study analyzed the secondary data from the Ethiopian Demographic Health Survey (EDHS), 2016, accessed from the Measure Evaluation Demography, Health Survey 2016 Ethiopia [10] which is freely available online [11] and contains information on a wide range of socioeconomic and demographic factors of the population nationwide. The country has nine regions and two administrative cities. The Ethiopian DHS 2016 utilized a two-stage sample design to select respondents for the study. In the first stage 645 enumeration areas (202 in urban areas and 443 in rural areas) were selected with probability proportional to size. Second stage involved selection of 28 households per cluster with an equal probability systematic selection from the newly formed household list. The EDHS 2016 has three parts: the household questionnaire, the woman's questionnaire, and the man's questionnaire. The data for child mortality and associated factors were taken from a woman's questionnaire. Data were collected by conducting face-to-face interviews with women who met the eligibility criteria (women aged 15–49 years).

Dependent variable: Status of stunting under five years old

Often in many public health studies, binary outcome is preferred as a response of interest for the sake of interpretation. Hence, our two responses were also studied as binary responses:

Stunted verses not stunted.

Y_{1i}= 1 if stunted (Z − score <−2)

0 if not stunted (Z − score >=-2) i.e. normal height for age

Independent variables: Months of breast feeding, sex of child, place of residence, education level, toilet facility, currently pregnant and child food nutrient. The description and coding of the independent variables are listed in **Table 1** below.

Table 1: Variables in the Study

No.	Variable Description	Code (If any)
1.	Breast feeding status	**0**=never breastfed; **1**=inconsistent
2.	Sex of child	**0**=female; **1**=male
3.	Age of child	**0**->29 months; **1**-<30 months
4.	Residence of child	**0**=urban; **1**=rural
5.	Level of education of Mother	**0**=no education; **1**=primary; **2**=secondary **3**=higher

6.	Use of toilet	**0**=unsafe; **1=**safe
7.	Pregnant status	**0**=no; **1**=yes
8.	Food nutrient status	**0**=no; **1**=yes
9.	Region	Addis Ababa = 0(ref), Tigray = 1, Afar = 2, Amhara = 3, Oromya = 4, Somali = 5, Benishangul-Gumuz = 6, SNNP = 7, Gambella = 8, Harari = 9, Dire Dawa = 10

2.2 Multilevel Logistic Regression Model

Two-Level Model

Multilevel models are statistical models which allow not only independent variable at any level of hierarchical structure but also at least one random effect above level one group [12]. A multilevel logistic regression model can account for lack of independence across levels of nested data (i.e., individuals nested within regions). Conventional logistic regression assumes that all experimental units are independent in the sense that any variable which affects occurrence of stunting has the same effect in all regions, but multilevel models are used to assess whether the effect of predictors vary from region to region.

In this study the basic data structure of the two-level logistic regression is a collection of N groups (regions) and within-group $j(j = 1,2,\dots,N)$, a random sample n_j of level-one units (children). The response variables, i.e., we let $Y_{ij} = 1$ if the ith under five children in jth region has stunting , and $Y_{ij} = 0$ otherwise; with probabilities, $P_{ij} = P(y_{ij} = 1|X_{ij}, u_j)$, is the probability of having stunting for child$i$ in region$j$ and $1 - P_{ij} = P(y_{ij} = 0|X_{ij}, u_j)$ is the probability of having no stunting for child$i$ in region$j$; where u_j is a random cluster effect and often assumed to be $N(0, \sigma_u^2)$. The standard assumption is that Y_{ij} has a Bernoulli distribution. Let P_{ij} be modeled using a logit link function. The two-level model is given by:

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{oj} + \sum_{l=1}^{k} \beta_{lj} x_{lij}\,; \qquad l = 1,2,\dots,k \qquad 1$$

Where $\beta_{0j} = \beta_0 + U_{0j}, \beta_{1j} = \beta_1 + U_{1j}, \dots, \beta_{kj} = \beta_k + U_{kj}$

The level-two model (1) can be rewritten as:

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_o + \sum_{l=1}^{k} \beta_l x_{lij} + U_{oj} + \sum_{l=1}^{k} U_{lj} x_{lij} \qquad 2$$

where $X_{ij} = (X_{1ij}, X_{2ij}, \dots, X_{kij})$ represent the covariates, $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ are regression coefficients, $U_{0j}, U_{1j}, \dots, U_{kj}$are the random effects of model parameter at level two. It is assumed that the $U_{0j}, U_{1j}, \dots, U_{kj}$ follow a normal distribution with mean zero and variance δ_u^2. Without U_{0j}, $U_{1j}, \dots, U_{kj}$, equation (2) can be considered as a single level logistic regression model. Therefore, conditional on $U_{0j}, U_{1j}, \dots, U_{kj}$, the y_{ij} can be assumed to be independently distributed as Bernoulli random variables [14].

Estimations of Between and Within Group Variance

The true variance between the group dependent probabilities, i.e. the population values of $Var(P_j)$, is given by:

$$\hat{\tau}^2 = S^2{}_{between} - \frac{S^2{}_{within}}{\tilde{n}} \qquad 7$$

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

For dichotomous outcome variables, the observed between group variance is closely related to the chi-square test statistic given in equation 5.

$S^2{}_{between} = \frac{\hat{p}(1-\hat{p})}{\tilde{n}\,(N-1)} X^2$ Where X^2 is given in equations (5).

The within group variance in case of a dichotomous outcome variable is a function of group averages which is given by:

$$S^2{}_{within} = \frac{1}{M-N} \sum_{j=1}^{N} n_j\, p_j (1 - p_j)$$

Multilevel logistic regression can be employed in the simplest case without explanatory variables (usually called empty model) and also with explanatory variables by allowing only the intercept term or both the intercept and the slopes (regression coefficients) to vary randomly. It mainly assumed that the varying coefficients have multivariate normal distribution [14].

2.2.1 The Empty Multilevel Logistic Regression Model

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

$$f(p_j) = \beta_o + U_{oj} \qquad 8$$

whereβ_o is the population average of the transformed probabilities and U_{oj}it is the random deviation from this average for group j. If f (p) is the logit function, then $f(p_j)$isjust the log-odds for group j. Thus, for the logit link function, the log-odds have a normal distribution in the population of groups, which is expressed by:

$$logit(p_j) = \beta_o + U_{oj} \qquad 9$$

For the deviations U_{oj}it is assumed that they are independent random variables with a normal distribution with mean zero and varianceσ_0^2.This model does not include a separate parameter for the level-one variance [14]. This is because the level-one residual variance of the dichotomous outcome variable follows directly from the success probability which is given by: $Var(\varepsilon_i) = P_j(1 - P_j)$

Denote by $\pi_0$the probability corresponding to the average value β_0, as defined by

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

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

$$\pi_0 = logistic(\beta_0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \qquad 10$$

Note that due to the non-linear nature of the logit link function, there is no a simple relation between the variance of probabilities and the variance of the deviationsU_{0j} [14]. An approximate variance of the probability given by:

$$var(P_j) \approx (\pi_0(1 - \pi_0))^2 \sigma_0^2 \qquad 11$$

Note that an estimate of population variance $var(P_j)$ can be obtained by replacing sample estimates of π_0 and σ_0^2. The resulting approximation can be compared with the nonparametric estimate, $\hat{\tau}^2$ which was given in equation (7).

2.2.2 The Random Intercept Model

In the random intercept model, the intercept is the only random effect meaning that the groups differ with respect to the average value of the response variable, but the relation between explanatory and response variables cannot differ between groups. We assume that there are variables which potentially explain the observed success and failure. These variables are denoted by $X_h, (h = 1,2, \ldots, k)$ with their values indicated by X_{hij}. Since some or all of those variables could be level one variables, the success probability is not necessarily the same for all individual in a given group [14]. Therefore, the success probability depends on the individual as well as the group, and is denoted by P_{ij}. the outcome variable is split into an expected value and residual as: $Y_{ij} = P_{ij} + R_{ij}$

The random intercept model expresses the log-odds, i.e. the logit of P_{ij}, as a sum of a linear function of the explanatory variables. That is,

$$\text{logit}(P_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{oj} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_k x_{kij}$$

$$= \beta_{oj} + \sum_{h=1}^{k} \beta_h x_{hij} \qquad 12$$

Where the intercept term β_{oj} is assumed to vary randomly and is given by the sum of an average intercept β_o and group-dependent deviations U_{oj}, that is $\beta_{oj} = \beta_o + U_{oj}$

As a result we have: $\text{logit}(P_{ij}) = \beta_o + \sum_{h=1}^{k} \beta_h x_{hij} + U_{oj}$ 13

Solving for P_{ij} we have: $P_{ij} = \frac{e^{\beta_o + \sum_{h=1}^{k} \beta_h X_{hij} + U_{oj}}}{1 + e^{\beta_o + \sum_{h=1}^{k} \beta_h X_{hij} + U_{oj}}}$ 14

Thus, a unit difference between the X_h values of two individuals in the same group is associated with a difference of β_h in their log-odds, or equivalently, a ratio of $\exp(\beta_h)$ in their odds. Equation (12) does not include a level-one residual because it is an equation for the probability P_{ij} rather than for the outcome Y_{ij}. Note that in the above equation $\beta_o + \sum_{h=1}^{k} \beta_h x_{hij}$ is the fixed part of the model. The remaining U_{oj} is called the random part of the model. It is assumed that the residual U_{oj} are mutually independent and normally distributed with mean zero and variance σ_o^2.

2.2.3 The Random Coefficient Multilevel Logistic Regression Model

In logistic regression analysis, linear models are constructed for the log-odds. The multilevel analogue, random coefficient logistic regression, is based on linear models for the log-odds that include random effects for the groups or other higher level units. Consider explanatory variables which are potential explanations for the observed outcomes. Denote these variables by $X_1, X_2, \dots, X_k$. The values of $X_h (h = 1, 2, \dots, k)$ are indicated in the usual way by X_{hij}. Since some or all of these variables could be level-one variables, the success probability is not necessarily the same for all individuals in a given group. Therefore, the success probability depends on the individual as well as the group, and is denoted by P_{ij}.

Now consider a model with group-specific regressions of logit of the success probability, $\text{logit}(P_{ij})$, on a single level one explanatory variable X,

$$\text{logit}(P_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{oj} + \beta_{1j} x_{1ij} \qquad 15$$

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

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

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

Substitution into (15) leads to the model

$$\text{logit}(P_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = (\beta_o + U_{oj}) + (\beta_1 + U_{1j}) x_{1ij}$$

$$= \beta_o + \beta_1 x_{1ij} + U_{oj} + U_{1j} x_{1ij} \qquad 16$$

There are two random group effects, the random intercept U_{oj} and the random slope U_{1j}. It is assumed that the level two residuals $U_{oj}$and U_{1j}have both zero mean given the value of the explanatory variableX. Thus, $\beta_1$is the average regression coefficient like β_o is the average intercept. The first part of equation 16 $\beta_o + \beta_1 x_{1ij}$ is called the fixed part of the model whereas the second part $U_{oj} + U_{1j}x_{1ij}$ is called the random part of the model.

The term $U_{oj} + U_{1j}x_{1ij}$ can be regarded as a random interaction between group and predictors (X). This model implies that the groups are characterized by two random effects: their intercept and their slope. These two groups effects $U_{oj}$and U_{1j} will not be independent. Further, it is assumed that, for different groups, the pairs of random effects (U_{oj}, U_{1j}) are independent and identically distributed. Thus, the variances and covariance of the level-two random effects (U_{oj}, U_{1j}) are denoted by:

$$Var(U_{oj}) = \sigma_{00} = \sigma_0{}^2$$

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

$$Cov(U_{oj}, U_{1j}) = \sigma_{01}$$

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

$$\mathrm{logit}(P_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{oj} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{kj}x_{kij} \qquad 17$$

Letting $\beta_{oj} = \beta_o + U_{oj}$ and $\beta_{hj} = \beta_h + U_{hj}$ where $h = 1, 2, \dots, k$, we have:

$$\mathrm{logit}(P_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_o + \sum_{h=1}^{k} \beta_h x_{ij} + U_{oj} + \sum_{h=1}^{k} U_{hj} x_{ij} \qquad 18$$

The first part$\beta_o + \sum_{h=1}^{k} \beta_h x_{ij}$is called the fixed part of the model, and the second part, $U_{oj} + \sum_{h=1}^{k} U_{hj} x_{ij}$ is called the random part of the model.The random variables or effects,$U_{0j}, U_{1j}, \dots, U_{kj}$are assumed to be independent between groups but may be correlated within groups. So the components of the vector $(U_{0j}, U_{1j}, \dots, U_{kj})$ are independently distributed as a multivariate normal distribution with zero mean vector and variances and co-variances matrix Ω given by:

$$\Omega = \begin{pmatrix} \sigma_0^2 & . & \cdots & . \\ \sigma_{01} & \sigma_1^2 & \cdots & . \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{0k} & \sigma_{1k} & \cdots & \sigma_k^2 \end{pmatrix}$$

2.2.4 Intra-class Correlation Coefficient (ICC)

The other fundamental reason for applying multilevel analysis is the existence of intra-class (intra-regional) correlation arising from similarity of incidence of stunting in the same region compared to those of different regions. The intra-class correlation coefficient (ICC) measures the proportion of variance in the outcome explained by the grouping structure. ICC can be calculated using an intercept-only model. This model can be derived from "Eq. (19)" by excluding all explanatory variables, which results in the following equation: (logit (p_j) = $\beta_{o+}U_{oj}$).The ICC is then calculated based on the following formula:

$$ICC = \frac{\delta_{uo}^2}{\delta_{uo}^2 + \delta_e^2} \qquad 19$$

whereδ_e^2 variance of individual (lower) level units

In multilevel logit model level one residual variance δ_e^2 =π2/3 ≈ 18 [14] this formula can be reformulated as:

$$ICC = \frac{\delta_{uo}^2}{\delta_{uo}^2 + 3.29} \qquad 20$$

For the purpose of model comparison study attempts the concept of maximum likelihood estimation via quadrature, AIC and BIC.