J Hum Ecol, 66(1-3): 1-11 (2019)
PRINT: ISSN 0970-9274 ONLINE: ISSN 2456-6608
DOI: 10.31901/24566608.2019/66.1-3.3142

The Use of Complex Survey Design Models to Identify Determinants of Malnutrition in Ethiopia

Ashenafi Argaw Yirga[1], Sileshi Fanta Melesse[1], Dawit Getnet Ayele[2] and Henry Mwambi[1]

[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa
[2] Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD 21205, USA

KEYWORDS Binary Outcome. Design Effect. Survey Logistic. Stunting. Underweight. Wasting

ABSTRACT Children's nutritional status has specific impact and health problems in childhood growth and adulthood. This paper attempts to identify the socio-economic, geographic and demographic determinants of malnutrition among under-five children in Ethiopia. The 2016 Ethiopian Demographic and Health Survey were used for this study. The generalized linear model was adopted for analysis. The Hosmer and Lemeshow test is used to test the goodness of fit of the logistic regression model. The findings and comparison of estimates using logistic regression model with and without complex survey design were presented and then comparison was made using design effects. The results revealed that model that considers the complex nature of the design performs better than model that do not take into account. This study suggests that improving the health status, stable work status and educational level of mothers consequently, can reduce malnourished children in Ethiopia.

INTRODUCTION

Child malnutrition is a very common public health problem in the world, especially in developing countries. The nutritional status of children under the age of five is an important outcome measure of children's health. This is because, the early days of child life is very important for future growth and development. Therefore, identifying factors that affect the nutritional status of under five children is very important for possible intervention activities. It can also assist policymakers to know and understand the areas that need considerable attention to enhance the planning and evaluation of health policies to prevent the child's death. For this reason, a national nutrition strategy and program has been developed and implemented by the government of Ethiopia (Ethiopian Demographic and Health Survey 2016). One of the objectives of the 2009 Ethiopian National Nutrition Strategy was to enhance good nutritional practices through health education, and treatment of micronutrients to the most vulnerable groups of the society, especially, for under five children and pregnant and lactating mother. However, the poor nutritional status of children and women has been a severe problem in Ethiopia. In 2016 Ethiopian Demographic and Health Survey (EDHS), children's nutritional status and health data were collected. In this nationally representative sample survey, measurements of children's weight and height were recorded. The purpose of these anthropometric measurements was to determine if children are growing normally. One of the important factors for child health is the child's weight or size at birth. This is the major indicator of the child's health which is related to childhood illnesses and survival. The major weight group with a higher risk of early death is for those children whose birth weight is less than 2.5 kilograms (CSA and ICF 2016).

Wasting, or low weight for height, is a strong predictor of mortality among children under five years of age. It is usually the result of acute significant food shortage and/or disease (Datar et al. 2013). According to UNICEF-progress for children 2007 report, there were 24 developing coun-

The present paper is extracted from the thesis title "Statistical models to study the BMI of under-five children in Ethiopia" of Ashenafi Argaw Yirga submitted at University of KwaZulu-Natal, School of Mathematics, Statistics and Computer Science, South Africa under the supervision of Dr. Sileshi Fanta Melesse, Prof. Henry Godwell Mwambi and Dr. Dawit Getnet Ayele.

Address for correspondence:
Dawit Getnet Ayele
Bloomberg School of Public Health,
The Johns Hopkins University,
Baltimore, MD 21205, USA
Cell: +1 (443) 554 6721
Email: dayele1@jhu.edu, ejigmul@yahoo.com

tries with wasting rates of ten percent or more, indicating a serious problem urgently requiring a response. The highest child malnutrition is found in the Sub-Saharan Africa countries. Ethiopia is among those countries with the highest rate of stunting in Sub-Saharan Africa. Universally, in 2011, eight percent of under five children were wasted (that is, weight for height below -2SD). This figure shows a eleven percent decrease compared to 1990 (De Onis et al. 2012). For the case of stunted children, the majority (90%) of children live in Africa and Asia. Moreover, sixteen percent of under five children were underweight (that is, weight for age below -2SD) in 2011. This is thirty six percent less compared to 1990. An estimated 159 million children under five years of age, or 23.8 percent, were stunting in 2016, 15.8 percent decrease from an estimated 255 million in 1990 worldwide (Achadi et al. 2016). For the case of Ethiopia, on an average twenty nine percent of children are underweight (low weight-for-age) and nine percent are severely underweight (CSA and ICF 2016). Even though the occurrence of stunting and underweight among children under five years of age worldwide has decreased since 1990, overall improvement is unsatisfactory and millions of children remain at risk (De Onis et al. 2012).

Determinants of Malnutrition of Children under Five Years in Rwanda was studied by in 2016. This study uses the 2010 Rwanda Demographic and Health Survey data. From the result, it was identified that the age of child, birth order, gender of child, birth weight, fever, mother's education level, mother's age at the birth, body mass index, mother's knowledge on nutrition, anemia, province, source of drinking water, multiple births, and wealth index have effects on malnutrition status (Habyarimana et al. 2016). The prevalence of malnutrition was studied in Sub-Saharan African Countries by Akombi et al. (2017). Their finding indicates that the prevalence is highest within countries in East Africa and West Africa. This was compared to the WHO Millennium development goals. In their study, they suggested that nutrition interventions have to be given attention in East Africa and West Africa.

Lara et al. (2017) studied the nutritional status of children in the Eastern Mediterranean Region (EMR). The result gives an important understanding of the nutrition situation of children. The result focused on two main parts of the EMR (Nasreddine et al. 2018). Therefore, from the result, it was possible to identify gaps and challenges in present nutritional assessment studies. The important points include the three-way problem of malnutrition in relation to underweight, nutrient inadequacies, and overweight/obesity.

Many research studies have shown how various socio-economic and demographic factors affect the malnutrition of children under the age of five. A recent study in Oromia region showed that thirty-five percent of non-pregnant women in this region had a BMI lower than 18.5, indicative of a high probability of getting underweighted children (Getahun et al. 2017). Underweight is commonly used as an indicator for malnutrition. It is influenced by the height and weight of a child/person and is thus a composite nature of stunting and wasting makes interpretation complex. The design of the Ethiopian Demographic and Health Survey was complex survey design. The sample was stratified and selected in two-stages with unequal sampling weights. Many studies used simple binary or ordinal logistic regression which does not include the complex nature of the sampling design. The purpose of this study was to identify the socio-economic, geographic and demographic determinants of malnutrition among under five children in Ethiopia by taking into account the complex nature of the sampling design.

METHODOLOGY

Variables of Interest

The response variable in this study is under five children's nutritional status (Healthy Weight Status and Malnourished), which is a binary outcome. The explanatory variables used in this study are: - child's age, sex of a child, weight of child at birth, mother's current age, mother's BMI, educational attainment of mother, mother's work status, religion, region, wealth index, place of residence, and current marital status. Significant two-way interaction effects were also considered. For this study, the 2016 Ethiopian Demographic and Health Survey were used. The socio-economic and demographic factors used in this study were supported by several researchers as most likely referred to as intermediate variables for the determinants of children nutritional status (Heeringa et al. 2010). The survey was carried out by the Central Statistical Agency of Ethiopia. A total of 16,650 households, 5,232 in urban and 11,418 in rural areas were selected in the survey. In the survey, 5,514 and 11,149 women aged 15-49 years were interviewed in urban and rural areas respectively and 14,195 men aged 15-59 years were interviewed of which 4,472 in urban and 9,723 in rural areas. The 2007 Ethiopia Population and Housing Census (PHC) was used as sampling frame for the 2016 EDHS. The 2016 EDHS sample was selected using a stratified, two-stage cluster design, and census enumeration area (EA). In the first stage, a total of 645 EAs (202 in urban areas and 443 in rural areas) were selected with probability proportional to EA size and with independent selection in each sampling stratum. In the second stage of selection, a fixed number of 28 households per cluster were selected with an equal probability systematic selection from the newly created household listing (CSA and ICF 2016).

The Statistical Model

Logistic regression is the appropriate regression model that can be used to study the relationship between a categorical or binary outcome variable and one or more explanatory variables. The logistic regression model is designed to describe the probability of an event, which is always some number between zero and one (Skinner et al. 1989). In the present study, such a probability refers to the chance of malnourishment of under five children.

For K explanatory variables and i=1, …, n individuals, the logit model is given by

$$logit(\pi) = log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

Where, π_ι is the probability that $y_i = 1$, β_0, is the intercept parameter, β_ι (i=1, 2,... k) is the slope parameters, and x_ι stands for explanatory variables. The expression on the left-hand side is the logit or log-odds.

Logit equation for (probability of the outcome of interest) is

$$\pi_i = P(Y_i = 1|X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}$$

Survey Logistic Regression

The survey method is an important methodology for survey data (Heeringa et al. 2010). Skinner et al. (1989) provided an important assessment of the expansion of methods for descriptive and analytical treatment of survey data.

When any sampling method other than simple random sampling is used, survey data analysis method has to be used. The survey analysis method is useful to include the design effect in the estimation of parameters and to adjust the standard errors of the estimates. If the sampling design is not included in the analysis, the standard errors will likely be underestimated, possibly leading to results that seem to be statistically significant, when in fact, they may not be significant. Therefore, this may lead us to biased estimates. Binary responses can be modelled through binary models that can provide a relationship between the probability of a response and a set of covariates. However, for data which does not come from simple random sampling, the standard logistic regression is not appropriate. According to Rao and Scott (1984), when the data comes from a complex survey design with stratification, clustering, and unequal weighting, the usual logistic regression estimates are not appropriate because the simple logistic regression does not account for clustered correlated observations. In these cases, specialized techniques that account for clustered correlated observations must be applied to produce the appropriate estimates and standard errors. The logistic regression model used to analyze data from complex sampling design is referred as survey logistic regression model. The theory behind survey logistic regression model is similar to the ordinary logistic regression model (Ayele et al. 2012). The difference between the two is that survey logistic regression accounts for the complexity of the sampling designs.

The model parameters in survey logistic regression model can be estimated by fitting a logistic regression model for binary outcome that takes into account the complex nature of the survey design. The survey logistic regression approach uses a modified maximum likelihood estimation called the 'pseudo-maximum likelihood estimation' (PMLE). It incorporates element weights in the estimating equation.

Let U={1,2,....N} be a finite population divided into h=1,2,....H strata, each stratum is further divided into j=1,2,...,n_h primary sampling units (PSU), which is constituted by i=1,2,...n_{hj} secondary sample units (SSU), comprehending n_{hji} elements. Assume, the observed data consists of n'_{hj} SSU chosen from n_h' PSU in the stratum h. Thus, the total number of the observation is to be:

$$n = \sum_{h=1}^{H} \sum_{j=1}^{n'_h} \sum_{i=1}^{n'_{hj}} n_{hji} .$$

Suppose that $\pi_{hji}=P(Y_{hji}=1|X_{hji})$, is the probability of an event in the sample, the sampling weight in each sampling unit is denoted by w_{hji}, for the *hji-th* unit. Thus, the survey logistic regression model is given by

$$logit(\pi_{hji}) = log\left\{\frac{\pi_{hji}}{1-\pi_{hji}}\right\} = X'_{hji}\beta .$$

Where, $Y_{hji,}$ $X_{hji,}$ and β are the categorical response variable, the covariate matrix, and the regression coefficients respectively.

The parameters β of the logistic regression model in the complex sampling design are estimated by the Pseudo maximum-likelihood method called weighted maximum likelihood that incorporates the sampling design and the different sampling weights in the estimation of β (Ayele et al. 2012; Hien and Hoa 2009).

The Pseudo-likelihood function for the contribution of a single observation in complex sampling design is given by

$$\pi_{hji}^{W_{hji}Y_{hji}} \left(1-\pi_{hji}\right)^{(1-W_{hji}Y_{hji})}$$

Thus, the Pseudo-likelihood function with weight W_{hji} for a set of n observation is given by

$$L(\boldsymbol{\beta}|W_{hji}Y_{hji}) = \prod_{h=1}^{H}\prod_{j=1}^{n'_h}\prod_{i=1}^{n'_{hj}} \pi_{hji}^{W_{hji}Y_{hji}}\left(1-\pi_{hji}\right)^{(1-W_{hji}Y_{hji})}$$

The main idea of this method is to define a function which approximates the likelihood function of the sampled finite population with a likelihood function formed by the observed sample and the known sampling weights (Ayele et al. 2012; Hien and Hoa 2009).

Estimates of parameters derived from complex sample survey data can be subject to substantial design effects, due to the stratification, clustering, and weighting associated with the design. Design effects (deff) are a measure of the relative effectiveness of the sample design, compared to a SRS. The square root of the deff, which is deft, serve as an inflation factor for the standard errors obtained using the complex sample. A deft=1 indicates no effect of sample design on standard error, deft>1 indicates sample design that inflates the standard error of the estimate, deft<1 indicates sample design increases efficiency that reduces standard error of the estimate.

RESULTS

The findings and comparison of results using the logistic regression model with and without complex survey design are presented in Tables 2 and 3. In this study, the probability of malnourishment of under five children is modelled as a function of the above mentioned explanatory variables.

The Likelihood Ratio test (LRT) was used to test the overall significance of the logistic regression model. The value of likelihood ratio statistic is 503.0579 with P-value <0.0001 which suggests the overall significance of the logistic regression model. The value of the score test is 504.8851 (P-value<0.0001) and the Wald test is 461.6973 (P-value<0.0001) which also supports the results obtained using the likelihood ratio test (Table 1). This shows that there is a significant contribution of independent variables in predicting the fitted logistic regression model. In other words, at least one of the parameters is significantly different from zero. The Hosmer and Lemeshow test for goodness of fit of this model is 13.9454 with P-value=0.0832, which shows that the model is a good fit to the data.

Table 1: Model evaluation for binary logistic regression

Model evaluation parameters	*Chi-square*	*D.F*	*P-value*
Overall significance			
Likelihood ratio test	503.0579	51	<.0001
Score test	504.8851	51	<.0001
Wald test	461.6973	51	<.0001
Goodness of fit test			
Hosmer and Lemeshow	13.9454	8	0.0832
Association of predicted probabilities and observed response			
Percent concordant	65.3	Somers 'D	0.307
Percent discordant	34.7	Gamma	0.307
Percent tied	0.0	Tau-a	0.098
Pairs	12897054	C	0.653

Another important aspect of the fitted logistic regression that needs to be checked is the validation of the model. The degree to which the predicted probabilities agree with the actual outcomes was expressed using a classification table with a cut-off point set at 0.5 (7). For better prediction power, the c-statistics has to be greater than 0.5. But it ranges from 0 (no association) to 1 (perfect association). The c-statistics is a measure of the predictive accuracy of a logistic regression model. In this particular study, the c-statistics is 0.653 (Table 1). This result shows that there is a moderate (65.3%) association between the predicted probabilities and the observed responses (actual probabilities). In addition, Table 1 shows that the concordant rate was 65.3 percent; this value tells us the agreement between the logistic regression model and the observed outcomes. The pairs indicates the number of pairs for 1 and 0. The Somer's D statistic is 0.307 suggesting that not all pairs are concordant. The Gamma statistic has a value of 0.307 which indicates a small positive association between variables. The estimated parameters of the fitted logistic model, the odds ratios and confidence interval for the odds ratios are presented in Table 2.

The predicted probability of malnourishment for a child as a function of the independent variables is estimated by

$$\hat{P}=\frac{e^{-3.1025+\sum_{i=1}^{\kappa}\beta_i\chi_i}}{1+e^{-3.1025+\sum_{i=1}^{\kappa}\beta_i\chi_i}}$$

Where $\hat{\beta}_i$'s are the estimated coefficients corresponding to x_i's, variables that have a significant effect on the response variable.

Table 2: Parameter estimates from the logistic regression model

Main effects	*Coeff. (β)*	*SE*	*OR*	*P-value*	*[95% C. I for OR]*
Intercept	-3.1025	0.6849		<.0001	
Current age of child	0.1974	0.1465	1.218	0.1779	(0.9306, 1.5052)
Sex of child (ref. Male) Female	-0.00628	0.0523	0.994	0.9044	(0.8915, 1.0965)
Weight of Child at Birth (ref. Small)					
Average	-0.3588	0.5212	0.698	0.4912	(-0.3235, 1.7195)
Large	-1.3413	0.5788	0.262	0.0205	(-0.8377, 1.3617)
Mother's current age	-0.00778	0.00414	0.992	0.0601	(0.9839, 1.0001)
Mother's BMI	0.1126	0.0329	1.119	0.0006	(1.0545, 1.1835)
Mother work status (ref. No) Yes	-0.1281	0.0607	0.880	0.0348	(0.7611, 0.9989)
Educational Attainment of Mother (ref. No education)					
Primary school	-0.6055	0.3329	0.546	0.0689	(-0.0925, 1.2125)
Higher	-0.1525	0.3713	1.165	0.5660	(0.4373, 1.8927)
Secondary school	-1.1423	0.2657	0.319	0.0021	(-0.2017, 0.8397)
Current marital status (ref. Married)					
Not married	-0.0320	0.1156	1.056	0.7820	(0.8294, 1.2826)
Religion (ref. Orthodox)					
Catholic	-0.2845	0.3836	0.7524	0.4583	(0.0006, 1.5043)
Muslim	0.1362	0.0949	1.1459	0.1511	(0.9599, 1.3319)
Other	0.3032	0.2143	1.3542	0.1571	(0.9342, 1.7743)
Protestant	0.2140	0.1102	1.2386	0.0521	(1.0226, 1.4546)
Region (ref. Oromia)					
Addis Ababa	1.3826	0.9283	3.985	0.1364	(2.1651, 5.8048)
Affar	1.6030	0.7888	4.968	0.0421	(3.4219, 6.5140)
Amhara	-0.0621	0.9079	0.939	0.9455	(-0.8405, 2.7185)
Benishangul	-0.00947	1.0408	0.991	0.9927	(-1.0489, 3.0309)
Dire Dawa	1.2040	0.8324	3.333	0.1480	(1.7015, 4.9645)
Gambela	-0.0578	0.9412	0.944	0.9510	(-0.9008, 2.7887)
Harari	0.8848	0.7846	2.422	0.2594	(0.8842, 3.9598)
SNNP	-0.3989	0.7423	0.671	0.5910	(-0.7839, 2.1259)
Somali	2.1369	0.7091	8.473	0.0026	(7.0832, 9.8628)
Tigray	0.8312	0.8834	2.296	0.3467	(0.5645, 4.0274)
Place of Residence (ref. Rural)					
Urban	-0.0178	0.1134	0.982	0.8756	(0.7597, 1.2042)
Wealth Index (ref. Poor)					
Middle	0.1176	0.0909	1.125	0.1959	(0.9469, 1.3031)
Richer	0.1803	0.0916	1.198	0.0490	(1.0185, 1.3775)
Significant Interaction Effects					
Age of child and Mother's BMI	-0.0247	0.00674	0.976	0.0002	(0.9628, 0.9892)
Age and Addis Ababa	0.2718	0.1022	1.312	0.0078	(1.1117, 1.5123)
Age and Amhara	-0.1482	0.0748	0.862	0.0474	(0.7154, 1.0086)
Age and Gambela	0.2328	0.0886	1.262	0.0086	(1.0884, 1.4356)
Mother's BMI and Addis Ababa	-0.0808	0.0398	0.922	0.0425	(0.8439, 1.0001)
Mother's BMI and Dire Dawa	-0.0782	0.0385	0.925	0.0423	(0.8495, 1.0004)
Mother's BMI and Somali	-0.1104	0.033	0.896	0.0008	(0.8313, 0.9606)
Mother's BMI and Large	0.0881	0.028	1.092	0.0016	(1.0371, 1.1468)

Figures 1-3 show the predicted probability of malnourishment of a child against significant interaction effects. The interaction effect plot between mother's BMI and region is presented in Figure 1. It is clearly seen that the predicted probability of malnourishment of a child among under five children increases as mother's BMI increase in Addis Ababa, Dire Dawa and Somali region.

The relationship between current age of a child and mother's BMI is presented in Figure 2. As the children's age increase, the probability of malnourishment of a child decreases across increasing mother's weight status.

The relationship between age of a child and regions (whereas region were Addis Ababa, Amhara, and Gambela) is presented in Figure 3. The prevalence of malnutrition was almost the same for under five children in Gambela region. As age increases, the prevalence of malnutrition decreased monotonically in Amhara region and Addis Ababa region.

The estimated parameters of the fitted survey logistic model (SAS PROC SURVEYLOGIS-

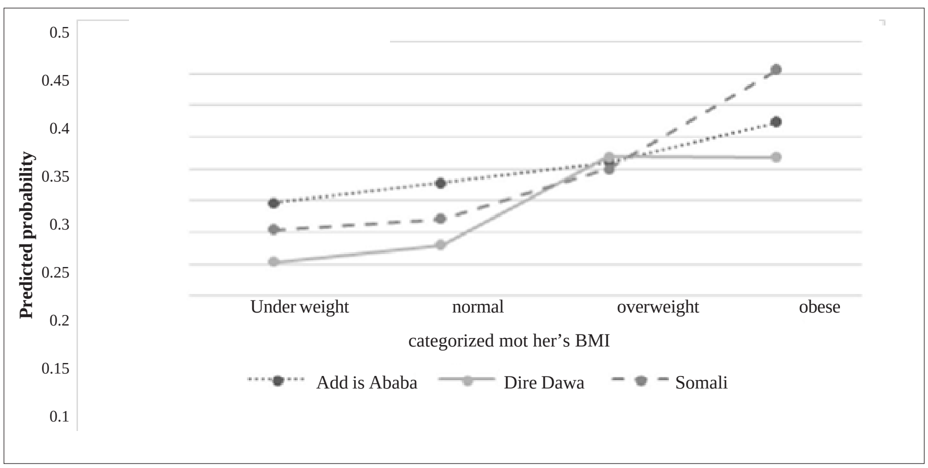

Fig.1. Predicted probability of malnutrition of under five children based on the effect of mother's BMI and region

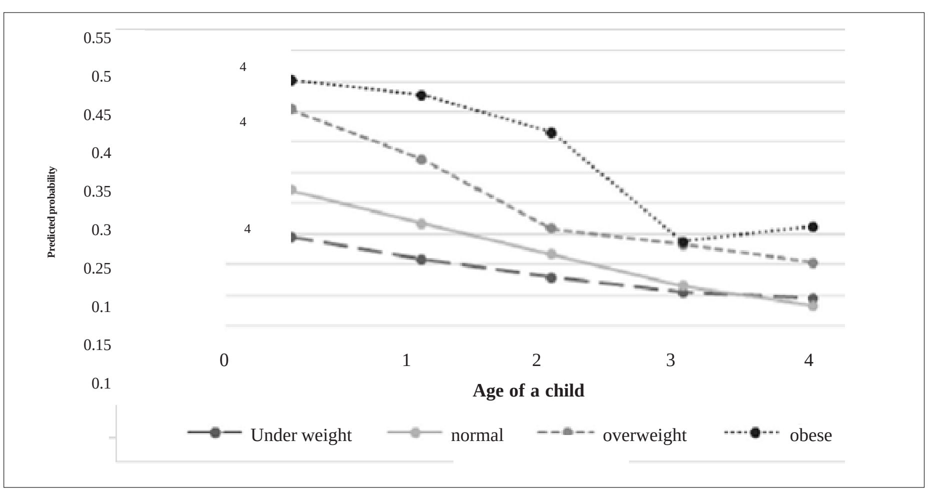

Fig. 2. Predicted probability of malnutrition of under five children based on the effect of mother's weight status and current age of a child

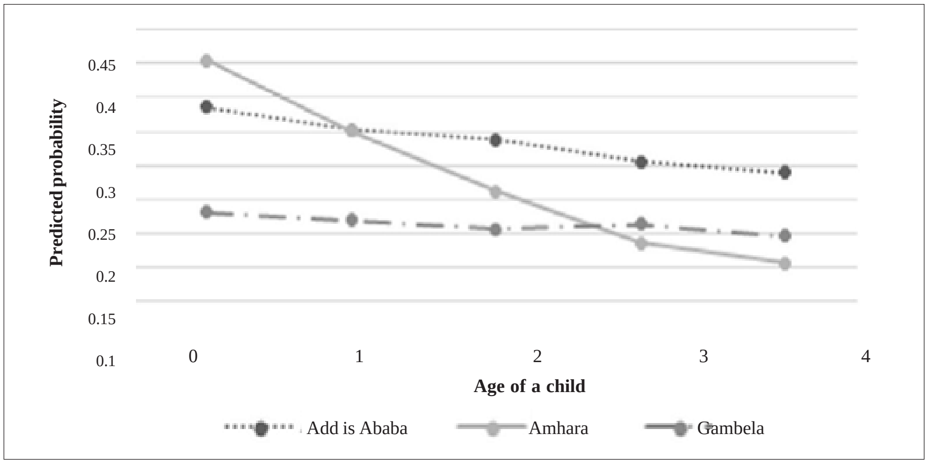

Fig. 3. Predicted probability of malnutrition of under five children based on the effect of region and age of a child

Table 3: Parameter estimates from the logistic regression model with complex survey design

Parameters	*Coeff. (β)*	*SE*	*OR*	*P-value*	*[95% C. I for OR]*
Intercept	-2.4657	0.7390		0.0009	
Current age of child	-0.00534	0.2375	0.995	0.9821	(0.5295, 1.4605)
Sex of child (ref. Male) Female	-0.0354	0.0766	0.965	0.9044	(0.8149, 1.1151)
Weight of Child at Birth (ref. Small)					
Average	-0.6758	0.8043	0.509	0.4012	(-1.0673, 2.0854)
Large	-1.0226	1.0041	0.359	0.3089	(-1.6090, 2.3270)
Mother's current age	-0.0139	0.00633	0.986	0.0283	(0.9735, 0.9984)
Mother's BMI	0.0991	0.0344	1.104	0.0041	(1.0366, 1.1714)
Mother work status (ref. No) Yes	-0.1808	0.0836	0.835	0.0310	(0.6711, 0.9989)
Educational Attainment of Mother (ref. No education)					
Primary school	-1.2956	0.4809	0.274	0.0073	(-0.6686, 1.2165)
Higher	0.1519	0.4892	1.164	0.6491	(0.2052, 2.1228)
Secondary school	-1.3937	0.3337	0.248	0.0045	(-0.4061, 0.9021)
Current Marital Status (ref. Married)					
Not married	-0.3395	0.1819	0.7121	0.0624	(0.3556, 1.0686)
Religion (ref. Orthodox)					
Catholic	0.0504	0.3360	1.052	0.8808	(0.3935, 1.7105)
Muslim	-0.1289	0.1227	0.879	0.2938	(0.6385, 1.1195)
Other	-0.0339	0.3305	0.967	0.9183	(0.3192, 1.6148)
Protestant	0.0508	0.1127	1.052	0.6523	(0.8311, 1.2728)
Region (ref. Oromia)					
Addis Ababa	1.4824	1.0436	4.403	0.1560	(2.3528, 6.4531)
Affar	-1.0380	1.4610	0.354	0.4777	(-2.5125, 3.2145)
Amhara	0.3899	1.0214	1.477	0.7028	(-0.5249, 3.4789)
Benishangul	0.3536	1.2129	1.424	0.7708	(-0.9533, 3.8012)
Dire Dawa	-2.1058	1.5642	0.122	0.1788	(-2.9438, 3.1878)
Gambela	-3.2258	3.2966	0.039	0.3282	(-6.4223, 6.5003)
Harari	-4.1726	1.7892	0.015	0.0200	(-3.4918, 3.5218)
SNNP	-0.3126	0.6985	0.732	0.6547	(-0.6370, 2.1010)
Somali	1.7657	0.9463	5.846	0.0626	(3.9912, 7.7007)
Tigray	-0.1939	1.0752	0.823	0.8569	(-1.2844, 2.9303)
Place of residence (ref. Rural) Urban	-0.1564	0.2249	0.855	0.4871	(0.4142, 1.2958)
Wealth Index (ref. Poor)					
Middle	0.1782	0.1238	1.195	0.1506	(0.9524, 1.4376)
Rich	0.2109	0.1256	1.235	0.0937	(0.9888, 1.4811)
Significant Interaction Effects					
Mother's BMI and Addis Ababa	-0.0862	0.0430	0.917	0.0456	(0.8327, 1.0013)
Mother's BMI and Somali	-0.1169	0.0415	0.889	0.005	(0.8077, 0.9703)

TIC), the odds ratios and confidence interval for the odds ratios are presented in Table 3. The interaction effect of mother's BMI and the Addis Ababa region was negatively associated with malnutrition of under five children (P-value=0.0456). The corresponding odds ratio was 0.917. This implies that the odds of malnourishment of under five children from Addis Ababa region decreased by 0.083 as compared to the odds of malnourishment of under five children from Oromia region who had mothers with average BMI. The effect of mother's BMI and the child from Somali region was also found to be negatively associated with malnutrition of under five children (P-value=0.005). The corresponding odds ratio was 0.889. This implies that the odds of malnourishment of under five children from Somali region decreased by 0.117 as compared to the odds of malnutrition of under five children from Oromia region who had mothers with the average BMI.

Based on the result from Table 3 the odds ratio for variables, which have significant effect on the probability of malnourishment of under five children are interpreted as follows: the odds ratio of 0.986 for mother's age indicates that for a one year increase in mother's age, the odds of having malnourished child decrease by 0.014. Furthermore, for one-unit increase in mother's BMI, the odds of having malnourished child will be multiplied by 1.104. The working status of mothers was found to be negatively associated with malnourishment of under five children as compared to mothers who were not working (P-value=0.031). The corresponding odds ratio was 0.835. In addition, the significant effects of covariates in terms of odds ratios in the logistic regression model with complex survey design can be interpreted in the same way as those in the logistic regression model without complex survey design.

Table 4 was constructed to compare standard error, confidence interval, and sample size obtained from PROC SURVEYLOGISTIC procedure (Table 3) with the ones obtained from PROC LOGISTIC procedure (Table 2) based on deff and deft. The explanatory variables presented were found to be significantly affecting malnourishment of under five children in Ethiopia. Both simple random sample (SRS) and complex survey design (design based) approaches are presented in Table 4.

Table 4 shows the deff and deft value for each significant effect in the study. Thus, the effect of mother's current age has the deff value of 2.3529 and deft value of 1.5339. The deft value equal to 1.5339, indicates that the sample standard error, and consequently the confidence interval are 1.5339 times bigger than they would be if the survey were based on simple random sampling. The effect of mother's BMI has deff=1.0901 and deft=1.0441. The deff value equal to 1.0901 indicates that the sample standard error, and consequently the confidence interval are 1.0901 times bigger than they would be if the survey were based on the same simple random sampling. The effect of currently working mothers have deff=1.8971 and deft=1.3773. The standard errors and confidence interval are 1.3773 times bigger than they would be for simple random sampling. The effect of mothers who had primary education level has deff=2.0868 and deft=1.4446. The deft value equal to 1.4446 indi-

Table 4: Comparison of estimates from the logistic regression model with and without CSD

	Model 1 (SRS)			*Model 2 (Design-based)*			*Study result*	
Parameters	*Coeff.*	*SE*	*P-value*	*Coeff.*	*SE*	*P-value*	*Deff*	*Deft*
Intercept	-3.1025	0.6849	<.0001	-2.4657	0.7390	0.0009	1.1642	1.0789
Mother's current age	-0.00778	0.00414	0.0601	-0.0139	0.00633	0.0283	2.3529	1.5339
Mother's BMI	0.1126	0.0329	0.0006	0.0991	0.0344	0.0041	1.0901	1.0441
Mother work status (ref. No)								
Yes	-0.1281	0.0607	0.0348	-0.1808	0.0836	0.0310	1.8971	1.3773
Educational Attainment of Mother (ref. No education)								
Primary school	-0.6055	0.3329	0.0689	-1.2956	0.4809	0.0073	2.0868	1.4446
Secondary school	-0.1525	0.3713	0.5660	-1.3937	0.4892	0.0045	1.7359	1.3175
Region (ref. Oromia)								
Harari	0.8848	0.7846	0.2594	-4.1726	1.7892	0.0200	5.2002	2.2804
Significant Interaction Effects								
Mother's BMI and Addis Ababa	-0.0808	0.0398	0.0425	-0.0862	0.0430	0.0456	1.1673	1.0804
Mother's BMI and Somali	-0.1104	0.033	0.0008	-0.1169	0.0415	0.005	1.5812	1.2575

cates that the sample standard error, and consequently the confidence interval are 1.4446 times bigger than they would be if the survey were based on the same simple random sampling. The effect of mothers who had secondary education level has deff=1.7359 and deft=1.3175. The deft value equal to 1.3175 indicates that the sample standard error, and consequently the confidence interval are 1.3175 times bigger than they would be if the survey were based on the same simple random sampling. The effect of a child from Harari region has deff=5.2002 and deft=2.2804. The deft value equal to 2.2804 indicates that the sample standard error, and consequently the confidence interval are 2.2804 times bigger than they would be if the survey were based on the same simple random sampling. The effect of mother's BMI depends on whether a child from Addis Ababa region has deff=1.1673 and deft=1.0804. The standard error and the confidence interval are 1.0804 times as bigger as they would be for the same simple random sampling. The effect of mother's BMI depends on whether a child from Somali region has deff=1.5812 and deft=1.2575. The deft value of 1.2575 indicates that the sample standard error, and consequently the confidence interval are 1.2575 times bigger than they would be if the survey were based on simple random sampling.

Furthermore, information on design effects should be used when we are planning to determine the sample size of the study. Once we have an estimated design effect, it is straightforward to adjust the required sample size; we need only to multiply the sample size needed under simple random sampling by the estimated design effect. For instance, deff=2.3529 (Table 4) indicates that for logistic regression studies with complex survey design, the sample size for mother's current age is 2.3529 times as large as would be needed under simple random sampling (logistic regression without complex survey design).

Models with smaller values of an information criterion are used to be considered preferable. However, the number of parameters estimated (k) in the complex model is higher than that of the conventional model. Since the researchers considered the complex nature of the design (weight, cluster and strata) in the binary model with complex survey design, AIC and BIC of the complex model are higher.

AIC and BIC can be viewed as measures that combine fit and complexity. The fit is measured negatively by -2 x ln (likelihood). Complexity is measured positively, either by 2 x k (AIC) or ln (N) x k (BIC) (see Akaike 1974; Schwarz 1978).

There are many different kinds of models that are available. Because of this, it is not inappropriate to use only a single criterion such as AIC (Melesse et al. 2016; Allison 2012). Other measures such as generalized cross-validation (GCV) are applied to validate the quality of the model. Nevertheless, it is important to make a choice by considering the nature of the models based on two different points. First, it is important to see how the model fits the observation. Secondly, the stability of the model in relation to estimation has to be considered (Tabachnick and Fidell 2007). These points were suggested by Hiruta and Asami (2016). Therefore, methods that take into account the complex nature of the design, perform better than those that do not take this into account. A complete review of types of methods to compare regression models is presented by different researchers (Tabachnick and Fidell 2007; Hiruta and Asami 2016).

DISCUSSION

In this study, the binary logistic regression model with and without complex survey design were presented to determine factors that affect malnourishment of under five children in Ethiopia. The socio-economic, geographic and demographic factors were used as explanatory variables. In addition, two-way interaction effects were included in the modelling process. Logistic regression also called a logit model is used to model dichotomous outcome variables. The Hosmer and Lemeshow test was used to test the goodness of fit of the binary logistic regression model. The goodness of fit tests indicate that the logistic model fits the data well. The parameter estimates obtained from both models were compared using design effects. The findings from these models show that the design effect values are above one. The result confirms that there was an underestimation of variance while using logistic regression. This is because it assumes data was sampled using simple random sampling. Since survey logistic regression accounts for the complexity of the survey design, it produced parameter estimates that are different from the estimates obtained when simple random sampling was assumed. However, in some cases, they were closer to one another. From the results, the researchers observed that the effect of mother's age, mother's BMI, uneducated mother, mother work status and Harari, Addis Ababa, Affar, Dire Dawa, Gambela and Somali region were found to have significant effect on the malnutrition of under five children in both models. Malnutrition includes a wide range of nutrient-related deficiencies and disorders whether it is due to dietary deficiency called under-nutrition, or to excess diet called over-nutrition (Lumley 2004). The risk of malnourishment of a child decreases with an increase in mother's age. The risk of malnourishment of a child increases with an increase in mother's BMI. The risk of malnourishment of a child born to a mother who was working was lower compared to the risk of malnourishment of a child born to a mother who was not working. The risk of malnourishment of a child born to a mother who had some educational level was lower compared to the risk of malnourishment of a child born to a mother who had no education. The risk of malnourishment also depends on region. For instance, the risk of malnourishment of a child in Somali region appeared to be higher compared to the risk of malnourishment of a child in Oromia region. According to 2016 EDHS report there has been a marked decline in the proportion of malnourishment of a child in Ethiopia. However, the findings of the researchers' study shows that regions specially Harari, Addis Ababa, Affar, Dire Dawa, Gambela and Somali region were more likely to experience malnourishment of under five children, which is a serious problem requiring a response among policy makers and need focus in order to enhance the planning and evaluation of health policies to prevent child death and to promote child health, diet and growth.

CONCLUSION

The findings of the researchers' study indicate that identifying the determinants of malnourishment is still an important issue among under-five children's in Ethiopia. The study also shows that not only education but also environmental and socio-economic factors were found to have significant effect on malnourishment of a child in Ethiopia. This study suggests that improving the nutritional status of the mother; consequently, improves the nutritional status of their children. Improving the stable work status of the mother will lead to an enhancement in mothers' economic status which will consequently improve the basic needs of their children. To change weight-related disorders, changes related to children, environmental and social intervention is required to promote and support weight-related change in mothers.

RECOMMENDATIONS

Based on researchers' study and other findings, they arrive at the following recommendations: the government of Ethiopia needs urgent implementation of poverty reduction strategies and program design to support the very poor families. Bringing rapid economic growth at national level is important to develop community-based interventions and could also serve as a long-term solution to the problem specially targeted to those regions of Harari, Addis Ababa, Affar, Dire Dawa, Gambela and Somali, which were highly affected by malnourishment of under five children. Children under five years of age needs a special health care. Therefore, policymakers need to focus on the influence of these significant factors to develop strategies that enhance normal or healthy weight status of children in Ethiopia.

LIST OF ABBREVIATIONS

BMI	: Body Mass Index
CDC	: Centers for Disease Control and Prevention
CSD	: Complex Survey Design
DEFF	: The Design Effect
DEFT	: The Square root of the DEFF
EDHS	: Ethiopian Demographic and Health Survey
EMR	: Eastern Mediterranean Region
SRS	: Simple random sampling
SD	: Standard Deviation

ACKNOWLEDGEMENTS

The researchers' thank, with deep appreciation, ORC Macro and Measure DHS for giving them access to the data file. This work was supported through the DELTAS Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and is supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107754/Z/15/Z- DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

REFERENCES

Achadi E, Ahuja A, Bendech MA, Bhutta ZA, De-Regil LM, Fanzo J, Fracassi P, Grummer-Strawn LM, Haddad LJ, Hawkes C et al. 2016. *Global Nutrition Report 2016: From Promise to Impact: Ending Malnutrition by 2030*. Washington DC: International Food Policy Research Institute.

Akaike H 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*,19(6): 716-723.

Akombi BJ, Agho KE, Merom D, Renzaho AM, Hall JJ 2017. Child malnutrition in Sub-Saharan Africa: A meta-analysis of demographic and health surveys (2006-2016). *PloS One,* 12(5): e0177338.

Allison PD 2012. *Logistic Regression using SAS: Theory and Application*. Cary, NC: SAS Institute 2012.

Ayele DG, Zewotir TT, Mwambi HG 2012. Prevalence and risk factors of malaria in Ethiopia. *Malaria Journal,* 11: 195.

CSA 2011. *Ethiopia Demographic and Health Survey 2011*. Addis Ababa, Ethiopia, and Calverton, Maryland, USA: CSA and ORC Macro.

CSA 2016. *Ethiopia Demographic and Health Survey 2016*. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF.

Datar A, Liu J, Linnemayr S, Stecher C 2013. The impact of natural disasters on child health and investments in rural India. *Social Science and Medicine,* 76: 83-91.

De Onis M, Brown D, Blossner M, Borghi E 2012. *Levels and Trends in Child Malnutrition. UNICEF-WHO-The World Bank Joint Child Malnutrition Estimates*. New York: UNICEF.

Getahun Z, Urga K, Ganebo T, Nigatu A 2017. Review of the status of malnutrition and trends in Ethiopia. *The Ethiopian Journal of Health Development (EJHD),* 15(2): 55-74.

Habyarimana F, Zewotir T, Ramroop S, Ayele D 2016. Spatial distribution of determinants of malnutrition of children under five years in Rwanda: Simultaneous measurement of three anthropometric indices. *Journal of Human Ecology,* 54: 138-149.

Heeringa SG, West BT, Berglund PA 2010. *Applied Survey Data Analysis*. Boca Raton, FL: CRC Press.

Hien NN, Hoa NN 2009. Nutritional status and determinants of malnutrition in children under three years of age in Nghean, Vietnam. *Pak J Nutr,* 8: 958-964.

Hiruta Y, Asami Y 2016. Method for Comparing Multiple Regression Models. *CSIS Discussion Paper No.* 141. Japan: Department of Urban Engineering, the University of Tokyo.

Hosmer DW, Lemeshow S 2000. *Applied Logistic Regression*. New York: John Wiley and Sons, Inc.

Lumley T 2004. Analysis of complex survey samples. *Journal of Statistical Software,* 9: 1-19.

Melesse S, Sobratee N, Workneh T 2016. Application of logistic regression statistical technique to evaluate tomato quality subjected to different pre-and post-harvest treatments. *Biological Agriculture and Horticulture,* 32: 277-287.

Nasreddine LM, Kassis AN, Ayoub JJ, Naja FA, Hwalla NC 2018. Nutritional status and dietary intakes of children amid the nutrition transition: The case of the Eastern Mediterranean region. *Nutrition Research,* 57: 12-27.

Rao JNK, Scott AJ 1984. On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics,* 12(1): 46-60.

Schwarz G 1978. Estimating the dimension of a model. *The Annals of Statistics,* 6: 461-464.

Skinner CJ, Holt D, Smith TF 1989. *Analysis of Complex Surveys*. UK: John Wiley and Sons.

Tabachnick BG, Fidell LS 2007. *Using Multivariate Statistics.* New York: Allyn and Bacon/Pearson Education.

Paper received for publication in April 2019
Paper accepted for publication in April 2019