Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2019;5(3): 226-235

Original Article

Application of Ordinal Logistic Regression Model to Nutritional Status of the Under-Five Children Indexed by Weight-for-Height

Anthony Ekpo[1]*, Waheed Babatunde Yahya[2]

[1]Department of Mathematics/Statistics/Computer Science,University of Agriculture, Makurdi, Nigeria.
[2]Department of Statistics,University of Ilorin, Ilorin, Nigeria.

ARTICLE INFO

Received 03.04.2019
Revised 21.06.2019
Accepted 23.07.2019
Published 01.08.2019

Key words:
Under-five;
Statistical surveillance;
Anthropometrics;
Bilateral edema;
Odd-ratio;
Nutritional status;
log-odds

ABSTRACT

Background and aim: In this paper, we present results regarding the outcomes of some anthropometric, epidemiological and demographic factors on the nutritional status of the under-five children which were categorized into three ordinal groups of Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM) and Global Acute Malnutrition (GAM) in Kazaure Local Government Area in Nigeria.
Methods: An ordinal logistic model that depicted the log-odds in favour of GAM (normal) child was fitted to the data based on surveillance indexed by Weight-For-Height (WFH).
Results: The results showed that the proportional odd of measuring the nutritional status of a child in a nutrition survey using the WFH index has the OR= 7.43 (95% CI, 4.717 to 11.705) times greater, with Wald $\chi^2(1)$ =74.81, p<0.001, hence a statistically significant effect.
Conclusion: Based on the results and summary of findings, it can be concluded that age is a major predictor of the nutrition status of a child in a nutritional study when the surveillance is based on WFH index unlike sex and measles that do not play a major role.

Introduction

Nutritional status is the best global indicator of well-being especially in children [7]. It is an integral component or reflection of the overall health of an individual and provides an indicator of the well-being of children living in a particular area. The nutritional status of a population is used to determine the magnitude of malnutrition that is prevailing at a particular time in a population. Nutritional status can be assessed using any of the following indicators: (i.) dietary intake (ii) biomarkers (iii.) anthropometry and (iv.) morbidity. Anthropometry has become a practical tool for evaluating the nutritional status of a population, particularly of children in developing countries [6]. Nutritional status is measured by the prevalence rates of stunting, wasting and underweight obtained through anthropometric measurements of children under five years. This research has focused on the three anthropometric

* . Corresponding Author: ekpo.anthony@uam.edu.ng

Please cite this article in press as Ekpo A, Yahya W.B. Application of Ordinal Logistic Regression Model to Nutritional Status of the Under-Five Children Indexed by Weight-for-Height. J Biostat Epidemiol. *2019; 5(3): 226-235*

indicators above singling out this assessment based on WFH [4].

2. Statement of the Research Problem

Due to recent global attention on nutritional surveillance and monitoring as a result of war, insurgencies, civil unrest and disease outbreaks around the world and particularly in Nigeria, this research therefore, seeks to use the Ordinal Logistics Regression (OLR) technique to model demographic factors (Age and Sex) of the child, Anthropometric factors (Height, Weight, and Mid-Upper-Arm-Circumference) and epidemiological factor (Measles and edema) with the outcome variable (nutritional status) categorized into; Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM) and Global Acute Malnutrition (GAM) respectively.

The main focus of this research work, therefore, is to establish the impact of some demographic, anthropometric and epidemiological variables on the malnutrition status of children in the Kazaure Local Government Area (LGA) of Jigawa state in northwestern Nigeria indexed by Weight-For-Height (WFH). To achieve this, a suitable logit model is fitted to model the impacts of these variables on the nutritional status of under-five children indexed by WFH in the Kasaure LGA. The graphical summary of the nutritional status of these children is also provided to aid the discussion of the results obtained.

2. Materials and Methods

2.1 Data Description:

The datasets used in this paper are primary data which emanated from a nutritional and mortality cross-sectional survey of children below five years of age (6-59 months) in Kazaure local government area of Jigawa State, North-Western Nigeria [4]. The cross-sectional survey used the Emergency Nutrition Assessment for Standardized Monitoring and Assessment of Relief and Transitions (ENA for SMART) software to gather the required data. In that survey, trained enumerators were used in the field to collect anthropometrics and other related variables from 604 children aged between 6-59 months with funding from Gunduma Health System, Jigawa State Ministry of Health with Technical Assistance from the researchers.

2.2 The Logit Regression Model

The logistic regression model is used to describe the relationship between the dichotomous response variable and a set of covariates [1]. The covariate variables may be continuous or discrete. Researchers often use logistic regression to estimate the effect of various covariates on some outcome of interest [2]. This basically assumed that in the logistic regression model, the log-odds of the outcome are linear functions of the covariates. That is the variables (X,Y,Z) are assumed to follow the model:

$$P(Y=1/X,Z)=H(\beta_0+\beta_1^T X+\beta_2^T Z)=H(\xi^T\chi). \quad (1)$$

Here the Y is a binary outcome while (X,Z) is a vector of covariates 8.

$$H(u)=\left[1+\exp(-u)\right]^{-1} \quad (2)$$

$$\xi=(\beta_0,\beta_1^T,\beta_2^Y)^T$$

is a vector of regression parameters for

$$\chi=(1,X^T,Z^T)^T \quad (3)$$

One can deduce from Ref. 3 that, the model for Logistic Regression Analysis (LRA) assumes that the outcome variable, *Y*, is categorical (e.g., dichotomous), but LRA does not model this outcome variable directly. Rather, it is based on probabilities associated with the values of *Y*. For simplicity, and because it is the case most commonly encountered in practice, we assume that *Y* is dichotomous, taking on values of 1 (i.e., the positive outcome, or success) and 0 (i.e., the negative outcome, or failure). In theory, the hypothetical, population proportion of cases for which, Y = 1 is defined as $p = P(Y=1)$. Then, the theoretical proportion of cases for which $Y = 0$ is $1 - p = P(Y = 0)$. In the absence of other information, we would estimate p by the sample proportion of cases for which $Y = 1$. However, in the regression context, it is assumed that there is a set of predictor variables, $X_1,\ldots,X_p$ that are related to Y and, therefore, provide additional information for predicting Y. For theoretical and mathematical reasons, LRA is based on a linear model for the natural logarithm of the odds (i.e., the log-odds) in favor of $Y = 1$ and is given by:

$$Log_e\left[\frac{P(Y=1\setminus X_1,\dots,X_P)}{1-P(Y=1\setminus X_1,\dots,X_P)}\right] = Log_e\left[\frac{\pi}{1-\pi}\right] = \alpha+\beta_1X_1+\dots+\beta_pX_p = \alpha+\sum_{j=1}^{p}\beta_jX_j \quad (4)$$

In particular, the inverse transformation is the logistic function of the form:

$$P = (Y=1/X_1,\dots,X_p) = \frac{e^{\alpha+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\alpha+\sum_{j=1}^{p}\beta_j x_j}} \quad (5)$$

The likelihood, L, of the sample data as the product, across all sampled cases, of the probabilities for success or failure:

$$L = \prod_{i=1}^{n} P(Y_i \setminus X_{i1},\dots,X_{ip}) = \prod_{i=1}^{n}\left[\left(\frac{e^{\alpha+\sum_{j=1}^{p}\beta_j x_j}}{1+e^{\alpha+\sum_{j=1}^{p}\beta_j x_j}}\right)^{\gamma} * \left(\frac{1}{1+e^{\alpha+\sum_{j=1}^{p}\beta_j x_j}}\right)^{1-\gamma}\right] \quad (6)$$

Analysis

The nutritional status was considered as the dependent/response variable while the three other selected factors were considered as the independent variables. The dependent variable was categorized into Severe Acute Malnutrition (SAM=1), Moderate Acute Malnutrition (MAM=2) and Global Acute Malnutrition(GAM=3) and demographic factors which included the Age of a child, which was collected in months with five categories (54-59 months =5, 42-53 months =4, 30-41 months =3, 18-29 months =2 and 6-17 months =1) and sex of the child was categorized into two; (Boys =1 and Girls =2), anthropometric factors Weight-for-Height (WFH) were categorized into (<-3 Z-score and/no edema = 1, <-2 and ≥-3 Z-score no edema = 2 and <-2 Z-score and/or edema =3), and finally under the anthropometric factor, and for the epidemiological factors, Measles and edema were considered, whilst measles was categorized into (Yes vaccinated = 3, Yes vaccinated, but no evidence =2 and Not vaccinated =1) presence of edema and no edema were classified into (Yes =1 and No=2) respectively.

Results

Data collation was done with ENA for SMART software and further analysis was carried out using the IBM-SPSS Statistics version 21 software, specifically the Ordinal Logistic Regression (OLR)–Polytomous Universal Model (PLUM) procedure [5].

Firstly, the independent and dependent variables selected for this study were subjected to these four assumptions below and were tested using SPSS to ascertain that:

The dependent variable is measured at the ordinal level.

one or more of the independent variables is continuous, ordinal or categorical (including dichotomous variables)

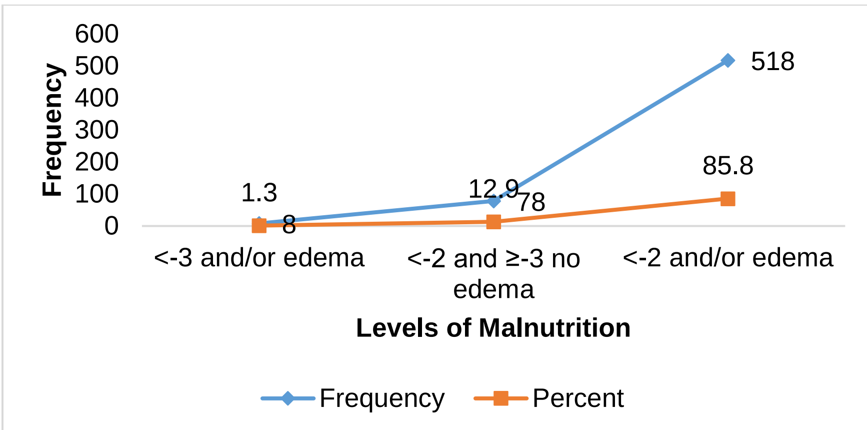
There are no issues of multicollinearity.

The variables satisfy the test for proportional odds

Table 1: Summary of Edema based on Weight-for-Height

Category based on WFH	Frequency	Percent	Cumulative Percent
<-3 and/or edema	8	1.3	1.3
<-2 and ≥-3 no edema	78	12.9	14.2
<-2 and/or edema	518	85.8	100.0
Total	604	100.0	

Figure 1.chart showing Edema Based on WFH

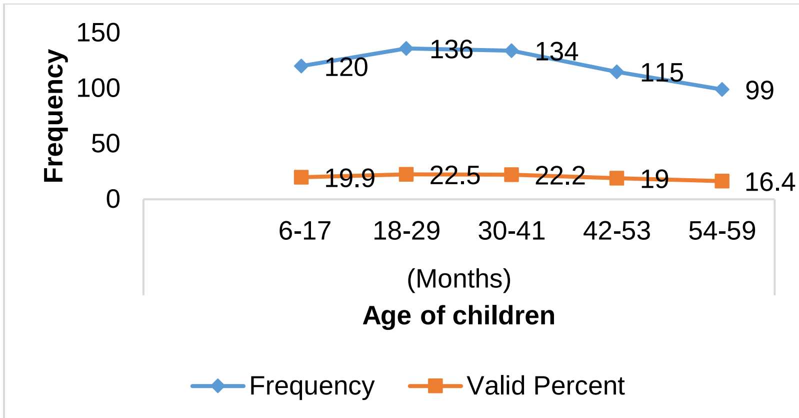


In table 1 and figure 1 above, we see the summary and chart of the number and percentage of edema based on WFH. These numbers look fine, but we would be concerned if one level had very few cases in it. We also see that all the 604 children surveyed were captured in the analysis. Only 1.3% of the children had edema based on WFH while 85.8% do not have also based on WFH.

Table 2: Summary Statistics of Age based on WFH

Age Interval	Frequency	Percent	Cumulative Percent
6-17	120	19.9	19.9
18-29	136	22.5	42.4
30-41	134	22.2	64.6
42-53	115	19.0	83.6
54-59	99	16.4	100.0
Total	604	100.0	

Figure 2: Chart showing age distribution of children Based on WFH

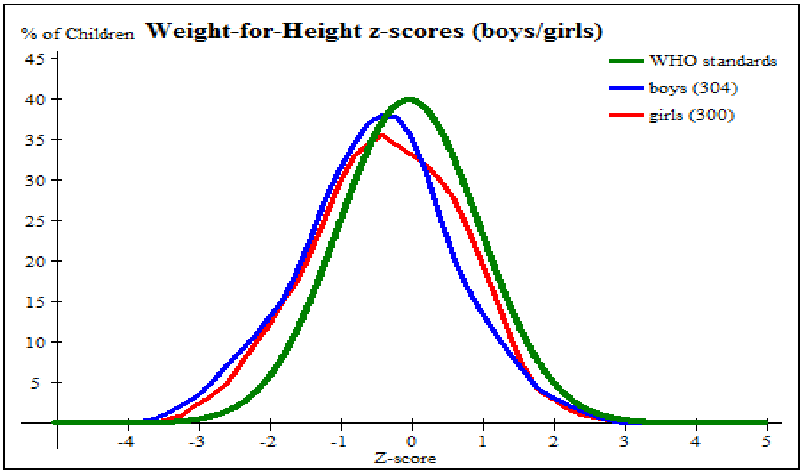


In table 2 and figure 2 above, we see the summary and chart of numbers and percentages of age distribution based on WFH. These spread look even, but we would be concerned if one level had extremely lower numbers in it. We also see that all the 604 children surveyed were captured in the analysis. Only 16.4% of the children are in the age bracket of 54-59 months while 22.5% are in the age bracket of 18-29 months.

Table 3: Descriptive Statistics of Sex based on WFH

Sex	Frequency	Valid Percent	Cumulative Percent
Male	304	50.2	50.2
Female	300	49.8	100.0
Total	604	100.0	

Fig 3: Plot of Z-score showing WFH of Boys and Girls



In table 3 and figure 3 above, we see the descriptive statistics and normal distribution plot of the numbers and percentages of sex distribution based on WFH. The skewness for the boys mimics the standard normalily more than that of the girls, but we would be concerned if any of them were negatively skewed. We also see that all the 604 children surveyed were captured in the plot.

Table 4. Summary of Measles Vaccination based on WFH

	Frequency	Percent	Cumulative Percent
No	18	3.0	3.0
Yes but no card	225	37.3	40.2
Yes	361	59.8	100.0
Total	604	100.0	

In table 4 above, we see the summary of measles vaccination based on WFH. These spread look good, especially because over 90% of the children were vaccinated for measles even though 37.3% could not show their cards. We also see that all the 604 children surveyed were captured in the summary. Only 3.0% of the children were not vaccinated.

Table 5. Model Fitting Information based on WFH

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	373.846			
Final	224.252	149.594	8	0.000

From table 5 above, which gives the 2-log likelihood for the intercept-only and final model, we could conclude that the model is good for prediction since significant level is less than 0.05.

Table 6. Goodness-of-Fit based on WFH

	Chi-Squaredf	Sig.

Pearson	81.885	1000.907
Deviance	93.963	1000.651

Table 6 also showed that both Pearson and Deviance chi-squares are not significant so we could get a good prediction for our model.

Table 7. Multicollinearity Coefficients based on WFH

Model	Collinearity Statistics	
	Tolerance	VIF
Age Cat	0.948	1.055
Sex	0.990	1.010
Measles	0.992	1.008
WFH	0.950	1.052

A check for multicollinearity from Table 7, showed that the tolerance for each of the variables are not less than 0.20 or 0.10 and the Variance Inflation Factor (VIF) is less than 5, hence there are no issues of multicollinearity.

Table 8: Summary Statistics of MUAC

Category based on MUAC	Frequency	Percent	Cumulative Percent
MUAC<110mm	1	0.2	0.2
MUAC 110-125mm	45	7.5	7.7
MUAC > 125mm	558	92.3	100.0
Total	604	100.0	

Table 9: Summary Statistics of Sex based on MUAC

Sex	Frequency	Percent	Cumulative Percent
Male	304	50.2	50.2
Female	300	49.8	100.0
Total	604	100.0	

Fig. 4: Cumulative distribution showing age distribution of children Based on MUAC

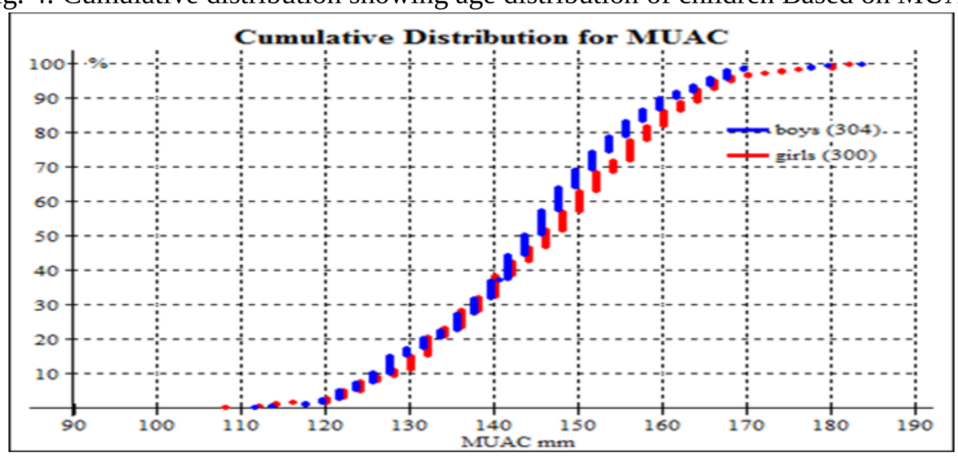


Table 10: Parameter estimates with proportional odds for WFH

Variables	***Estimate***	***Std. error***	***Wald Chi-Squ***	***DF***	***P-value***	***Lower Bound***	***Upper Bound***	***Odd Ratio***	***Lower***	***Upper***
[Nut_Status = 1]	1.794	0.715	6.284	1	0.012	0.391	3.196	6.011	1.479	24.432
[Nut_Status = 2]	3.643	0.74	24.222	1	<0.001	2.192	5.094	38.207	8.955	163.006
WFH	2.006	0.232	74.81	1	0.000	1.551	2.46	***7.43***	4.717	11.705
[Age_Cat=1]	-1.274	0.39	10.687	1	0.001	-2.038	-0.51	0.28	0.13	0.6
[Age_Cat=2]	-2.245	0.375	35.754	1	0.000	-2.981	-1.509	0.106	0.051	0.221
[Age_Cat=3]	-1.689	0.383	19.427	1	0.000	-2.44	-0.938	0.185	0.087	0.392
[Age_Cat=4]	-1.04	0.404	6.635	1	0.001	-1.831	-0.249	0.354	0.16	0.78
[Age_Cat=5]	0			0*				1		
[Sex=1]	-0.089	0.184	0.236	1	0.627	-0.449	0.271	0.915	0.638	1.311
[Sex=2]	0			0*				1		
[Measles=1]	0.286	0.547	0.273	1	0.601	-0.786	1.358	1.331	0.456	3.888
[Measles=2]	0.224	0.193	1.347	1	0.246	-0.155	0.603	1.252	0.857	1.828
[Measles=3]	0			0*				1		

From table 10 above, we can see that the odds of using the Weight-for-Height (WFH) in classifying the Nutritional Status of a child to be either GAM versus combined MAM and SAM is 7.43 (95% CI, 4.717 to 11.705) times greater, likewise, the odds of the combined MAM and GAM versus SAM is 7.43 times greater with Wald $\chi^2(1)$ =74.81, p<0.001, hence a statistically significant effect.

An increase in the age of a child within the age interval (6-17 months) was associated with an increase in the odds of classifying the Nutritional status of the child to be either GAM versus MAM and GAM with an odds ratio of 0.28 (95% CI, 0.13 to 0.6) likewise, the odds of the combined MAM and GAM versus SAM is 0.28 greater with Wald $\chi^2(1)$ =10.687, p = 0.001, hence a statistically significant effect.

An increase in the age of a child within the age interval (18-29 months) was associated with an increase in the odds of classifying the Nutritional status of the child to be either GAM versus combined MAM and GAM with an odds ratio of 0.106 (95% CI, 0.051 to 0.221) likewise, the odds of the combined MAM and GAM versus SAM is 0.106 greater with Wald $\chi^2(1)$ =35.754, p < 0.001, hence a statistically significant effect.

An increase in the age of a child within the age interval (30-41 months) was associated with an increase in the odds of classifying the Nutritional status of the child to be either GAM versus MAM and SAM with an odds ratio of 0.185 (95% CI, 0.087 to 0.392) likewise, the odds of the combined MAM and GAM versus SAM is 0.185 times greater with Wald $\chi^2(1)$ =19.427, p < 0.001, hence a statistically significant effect.

Similarly, an increase in the age of a child within the age interval (42-53 months) was associated with an increase in the odds of classifying the Nutritional status of the child to be either GAM versus combined MAM and GAM with an odds ratio of 0.354 (95% CI, 0.16 to 0.78) likewise, the odds of the combined MAM and GAM versus SAM is 0.354 times greater with Wald $\chi^2(1)$ =6.635, p = 0.01, hence a statistically significant effect.

Again the odds of the sex of a child being male as a consideration of his nutritional status is 1.915 times that of the females when WFH is used as the index for measuring the prevalence of malnutrition.

The odds of a child who was not vaccinated for measles to be considered as GAM versus MAM and GAM is 1.331 times greater (95% CI, 0.456

to 3.888) that of being vaccinated and vaccinated with no evidence likewise, the odds of the combined MAM and GAM versus SAM with Wald $\chi^2(1)$=0.273, p = 0.601, hence a not statistically significant effect. Similarly, the odds of a child whose caregiver claim was vaccinated but with no evidence to be considered to be GAM versus MAM and GAM is 1.252 (95% CI, 0.857 to 1.828) times that of being vaccinated and not being vaccinated likewise, the odds of the combined MAM and GAM versus SAM with Wald $\chi^2(1)$=1.347, p = 0.246, hence a not statistically significant effect.

5.Summary of Findings

In fitting an illustrative model with the data from under-five nutritional survey and based on the illustrative examples of Ref. 3 and the SPSS PLUM result on table 6 the following illustrative models are given:

$$\log_e\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 5.437 + 2.006*WFH - 6.248*AGE - 0.089*SEX + 0.51*MEASLES$$

To find out whether the number of independent variables considered here predict the ordinal dependent variable, "nutritional status" we observed that the reported standard error for the estimates in table 8, the standard error for the estimate, 0.375 as given for age category 2 alone, with a statistical significance which can be assessed by the Wald chi-squares 35.754 at conventional levels (the empirical 2-tailed p-values reported to be 0.000).

To determine whether WFH have a statistically significant effect on the dependent variable, we observed that at the significance (p-values) on tables 10 age at category 1, 2, 3 and 4 were statistically significant on the dependent variable with Wald chi-squares 10.687, 35.754, 19.427 and 6.635 respectively, and were significant at conventional levels (the empirical 2-tailed p-values reported to be 0.001, 0.000, 0.000 and 0.001) respectively, which also supports the fact that a child's age is a useful predictor of children's nutritional status.

In classifying the proportional odds of being in either of the nutritional status, we found out that cumulatively, an increase in the age of a child was associated with the increase in the odds of considering his nutritional status to be either severely malnourished versus moderately malnourished and globally malnourished (normal) was 1.925 times greater when WFH was used as the index for measuring the prevalence of malnutrition, likewise, the odds of the combined moderately malnourished and globally malnourished (normal) versus severely malnourished.

Finally, the odds of a child who was not vaccinated for measles to be considered as either severely malnourished, moderately malnourished or globally malnourished is 3.583 times greater than that of not being vaccinated and vaccinated but with no evidence when WFH is used as the index for measuring the prevalence of malnutrition.

6.Conclusion and Recommendation

Based on the results and summary of findings, it can be concluded that age is a major predictor of the nutrition status of a child in a nutritional study when the surveillance is based on WFH index unlike sex and measles that do not play a major role. Similarly, measuring malnutrition prevalence using the WFH index happens to give highest odds when compared with other indices.

Given the results and findings in this study, it can be recommended that age is an indicator variable and should always be included as a measure in child's nutritional status. Sex does not play a major role in determining nutritional status; it can only give an idea of the summary statistics of gender considered in the study. Measles as an epidemiological state did not play a major role in classifying a child's malnutrition status hence it could be excluded from the list of variables, rather variables that have to do with access to potable water, access to household food for children and other socio-economic variables should be considered. Since WFH index happens to have the highest odds, it means that using it in a nutrition survey to categorize nutritional status would give much more reliable prevalence of SAM and GAM when compared with other indices.

References

[1] K. A. Adeleke, and A. A. Adepoju, Ordinal Logistic Regression Model:
An Application to Pregnancy Outcomes, *J. Maths and Stats.* **6** (3) (2010) 279-285.

[2] R.A. Adeleke, et al., Application of Ordinal Logistic to pregnancy outcomes,
J. Natural Sci. Research. **2**(4) (2012) 1-7.

[3] C. M. Dayton, Logistic Regression Analysis, (ResearchGate Publications, 1992. pp. 1- 9).

[4] A. Ekpo and E. F. Udoumoh, Cluster Surveillance of Malnutrition Prevalence and Hunger Pattern in Kazaure Local Government Area of Jigawa State, Nigeria (2010- 2012) in *Proc. 59th Int. Statistical Inst. World Statistics Congress*, (Hong Kong, China. 2013). pp. 4679-4681.

[5] Enhanced SPSS guides in Laerd Statistics-www.statistics.laerd.com

[6] R. Hakeem, A. H. Shaikh and F. Asar Assessment of Linear Growth of Affluent Urban Pakistani Adolescents According to CDC 2000 References, *Annals of Human Biology*. **31**(3) (2004) 282-291.

[7] Monika, De Onis, Blössner, Mercedes. Malnutrition: quantifying the health impact at national and local levels, *Environmental Burden of Disease Series*, Series 12. (World Health Organization, Geneva, 2005).

[8] R. Ante and L. Ante, Using Ordinal Logistic Regression to Predict Terms of Payment in Ship building Industry, in *Proc. 58th Int. Statistical Inst. World Statistical Congress, (*Dublin 2011). pp. 5818-5821.