

# Bayesian Modelling of Growth Retardation among Children Under-Five Years Old in Ethiopia

Kasahun Takele

Department of Statistics, Haramaya University

[kastake0@gmail.com](mailto:kastake0@gmail.com)

## ABSTRACT

*Malnutrition among children under age five is the major public health problem in the developing world particularly in Ethiopia. The aim of this study was then to determine statistically the determinants of children malnutrition, using 2011 DHS data. The overall prevalence of stunting among children in Ethiopia was 43.3%. Bayesian Semi-parametric regression model was used to flexibly model the effects of selected socioeconomic, demographic, health and environmental covariates. Inference was made using Bayesian approach with Markov chain Monte Carlo (MCMC) techniques. It was found that the covariates sex of child, preceding birth interval, birth order of child, place of residence, region, mother's education level, husband educational level, toilet facility, number of household members, household economic status, diarrhea and fever were the most important determinants of children nutritional status in Ethiopia. The effect of child Age, mother's age at child birth and mothers body mass index were also explored non-parametrically as determinants of children nutritional status. It is suggested that for reducing childhood malnutrition, due emphasis should be given in improving the knowledge and practice of parents on appropriate young child feeding practice and frequent growth monitoring together with appropriate and timely interventions.*

**Keywords:** *Undernutrition, Child, Bayesian Semiparametric regression model, MCMC*

## **1. INTRODUCTION**

The importance of nutrition on early-childhood development outcomes has gained international awareness. Strong evidence shows that nutritional failure during pregnancy and in the first two years of life lead to lower human capital endowments, negatively affecting physical strength and cognitive ability in adults. This contributes directly reduced earnings potential of individuals and damages national economic growth and competitiveness potential in a globalized world (World Bank, 2007).

Nutritional status during childhood has consequences in childhood until adulthood. Deficiencies in nutrients or imbalances between them can have dire long-term effects for the individual (Kibel et al., 2007). Thus, measuring the child's nutritional status is important because of both the long-term and short-term effects on the health, educational and the cognitive abilities of the child. There are also severe consequences and effects to the child's ability to function as a healthy, productive and self-supporting community member in the long-term, which is another reason for concern, and as such the study further wishes to add to an understanding on how to contribute to the betterment of society.

Measures of child malnutrition are based on height-for-age, weight-for-age, and weight-for-height. Each of these indices provides somewhat different information about the nutritional status of child. The height-for-age index measures linear growth retardation among children, primarily reflecting chronic malnutrition. The weight-for-height index measures body mass in relation to body height, primarily reflecting acute malnutrition. Weight-for-age reflects both chronic and acute malnutrition (WHO, 1995; Maleta, 2006).

Nutritional status of children in Ethiopia is among the worst in the world. For example, chronic malnutrition in Ethiopia is worst than other SSA countries: about one in two children (51 percent) are moderately to severely stunted, of which slightly more than one in four children (26 percent) are severely stunted . Thus, high malnutrition rates in Ethiopia pose a significant obstacle to achieve better child health outcomes (NNS, 2009).

The implication of this high prevalence of child malnutrition is that a good knowledge regarding the major factors that contribute to the problem is essential in order to avoid its adverse consequences. The causes and determinants of child malnutrition are complex, interrelated, and multidimensional. In the literature, since mothers are the main providers of primary care to their children, understanding the contribution of maternal characteristics on child nutrition has been identified as a key towards addressing the problem of child malnutrition. In Ethiopia reason behind malnutrition is still to be known from recent data available. Therefore, the study was to model the various possible factors and their contribution for the current high prevalence of malnutrition problems using the Bayesian semi-parametric regression model.

## **2. DATA SOURCE AND METHODOLOGY**

### **2.1. Data Source**

This research used the 2011 Ethiopian Demographic and Health Survey. The survey drew a representative sample of women of reproductive age, by administering a questionnaire and making an anthropometric assessment of women and their children that were born within the previous five years. The analysis presented in this study on nutritional status was based on the 7739 children aged less than 60 months with complete anthropometric measurements.

## **2.2. Variables used in the study**

As verified in the background, socio-economic, demographic, health and environmental characteristics are considered as the most important determinants of child nutritional status. In our application on children nutritional status, stunting is used which is the response variable. Z-score (in a standardized form) was used as a continuous variable to maximize the amount of information available in the data set. The explanatory variables which might determine nutritional status of child were socio-economic, demographic, health and environmental factors. These factors include the sex of child, age of child, preceding birth interval, birth order of child, mother's age at child birth, place of residence, region, mother's education level, husband educational level, mother's work status, number of household members, household economic status, mothers body mass index, diarrhea, fever, water supply and toilet facility.

## **2.3. Methods of Statistical Analysis**

Bayesian methods have become popular in modern statistical analysis and are being applied to a broad spectrum of scientific fields and research areas. Bayesian data analysis involves inferences from data using probability models for quantities we observe and for quantities about which we wish to learn or in other words analyzing statistical models with the incorporation of prior knowledge about the model or model parameters.

The statistical analysis in this research is based on Bayesian approaches which allow a flexible framework for realistically complex models. These approaches allow us to analyze usual linear effects of categorical covariates and nonlinear effects of continuous covariates within a unified semi-parametric Bayesian framework for modeling and inference. This study used generalized

additive models to simultaneously incorporate the usual linear effects as well as nonlinear effects of continuous covariates within a semi-parametric Bayesian approach. The inference we make is fully Bayesian and uses recent Markov Chain Monte Carlo (MCMC) simulation techniques for drawing random samples from the posterior.

### 2.3.1. Bayesian Semi-parametric Regression Models

The assumption of a parametric linear predictor for assessing the influence of covariate effects on responses seems to be rigid and restrictive in practical application situation and also in many real statistically complex situation since their forms cannot be predetermined a priori. In this application to childhood under-nutrition and in many other regression situations, we are facing the problem for the continuous covariates in the data set; the assumption of a strictly linear effect on the response  $Y$  may not be appropriate as suggested in Khalid (2007), Mohammed (2008) and Kandala (2008).

Traditionally, the effect of the covariates on the response is modelled by a linear predictor as:

$$\eta_i = X_i' \beta + W_i' \gamma \quad (1)$$

Where:

$X_i = (X_{i1}, \dots, X_{ik})$  is a vector of continuous covariates,

$\beta = (\beta_0, \beta_1, \dots, \beta_p)$  is a vector of regression coefficients for the continuous covariates.

$W_i = (w_{i1}, \dots, w_{ik})$  is a vector of categorical covariates.

$\gamma = (\gamma_1, \dots, \gamma_k)$  is a vector of regression coefficients for the categorical covariates.

In the Bayesian parametric regression model, the parameter vectors  $\beta$  and  $\gamma$  one routinely assume diffuse priors  $p(\gamma) \propto \text{const}$  and  $p(\beta) \propto \text{const}$ . A possible alternative would be to work with a multivariate Gaussian distribution  $\gamma \sim N(\gamma_0, \Sigma_{\gamma_0})$  and  $\beta \sim N(\beta_0, \Sigma_{\beta_0})$ . However, since in most cases a non-informative prior is desired, it is sufficient to work with diffuse priors.

In this study, the continuous covariates child's age (Cage), the mother's age at birth (Mage), and the mother's Body Mass Index (BMI) are assumed to have non-linear effects on child nutritional status. Hence, it is necessary to seek for a more flexible approach for estimating the continuous covariates by relaxing the parametric linear assumptions, by considering their true functional forms. This can be done using an approach referred to as non-parametric regression model. Non-parametric regression analysis is regression without an assumption of linearity. The scope of non-parametric regression is very broad, ranging from "smoothing" the relationship between two variables in a scatter plot to multiple-regression analysis and generalized regression models (for example, logistic non-parametric regression for a binary response variable).

To specify a non-parametric regression model, an appropriate function that contains the unknown regression function needs to be chosen. This choice is usually motivated by smoothness properties, which the regression function can be assumed to possess.

The semi-parametric regression model is obtained by extending model () as follows:

$$\eta_i = f_1(x_{i1}) + \dots + f_k(x_{ik}) + w_i' \gamma \quad (2)$$

$$i = 1, 2, \dots, n$$

$$(k = 3)$$

Where,  $f_1, \dots, f_k$  are smooth functions of the continuous covariates.

### 2.3.2. Prior Distributions

Existing evidence about the parameters of interest may be available from earlier studies or from experts' opinions and can be formalized into what is called prior distribution of the parameter of interest. A prior distribution can be non-informative, informative, or very informative. Non-informative prior distributions are used in cases in which no extra-sample information is available on the value of the parameters of interest (Clark et al., 2002 and Mila et al., 2003). In statistical terms, this lack of knowledge is represented with a distribution that attributes, approximately, the same probability to each possible parameter value.

In model (2), the parameters of interest  $f_j, j = 1, \dots, p$  and parameters  $\gamma_i$  as well as the variance parameter ( $\tau^2$ ) are considered as random variables and have to be supplemented with appropriate prior assumptions. In the absence of any prior knowledge we assume independent diffuse priors  $\gamma_j \propto \text{const}, j = 1, \dots, r$  for the parameters of fixed effects. Another common choice is highly dispersed Gaussian priors.

Several alternatives are available as smoothness priors for the unknown functions  $f_j(x_j)$ . Among the others, random walk priors (Fahrmeir and Lang (2001), Bayesian Penalized-Splines (Fahrmeir, Kneib and Lang (2004), Bayesian smoothing splines (Hastie and Tibshirani, 2000) are the most commonly used. In this study, the Bayesian smoothing spline was used by taking cubic P-spline with second order random walk priors (Kandala, 2010; Mohammed, 2008).

Suppose that  $f = (f(x_1), \dots, f(x_n))'$  is the vector of corresponding function evaluations at observed values of  $X$ .

Then, the prior for  $f$  is

$$\left[ f \mid \tau^2 \right] \propto \exp\left( -\frac{1}{2\tau^2} \mathbf{f}' K \mathbf{f} \right) \quad (3)$$

Where,  $K$  is a penalty matrix that penalizes too abrupt jumps between neighboring parameters. In most cases,  $K$  will be rank deficient; therefore the prior for  $f$  would be improper. This implies that  $(f/\tau^2)$  follows a partially improper Gaussian prior  $f/\tau^2 \sim N(0, \tau^2 K^-)$

where  $K^-$  is a generalized inverse of a band-diagonal precision or penalty matrix  $K$ . it is possible to express the vector of function evaluations  $f_j = (f_j(x_{1j}), \dots, f_j(x_{nj}))'$  of a nonlinear effect as the matrix product of a design matrix  $X_j$  and a vector of regression coefficients  $\beta_j$ ,

$$f_j = X_j' \beta_j$$

Brezger and Lang (2006) also suggest a general structure of the priors for  $\beta_j$  as:

$$p(\beta_j \mid \tau^2) \propto \frac{1}{(\tau^2)^{\text{rank}(K_j)/2}} \exp\left(-\frac{1}{2\tau^2} \beta_j' K_j \beta_j\right) \quad (4)$$

where  $K_j$  is a penalty matrix that shrinks parameters towards zero or penalizes too abrupt jumps between neighboring parameters. In most cases,  $K_j$  will be rank deficient and, therefore, the prior for  $\beta_j$  is partially improper. The penalty matrix is of the form  $K = D'D$ , where  $D$  is a first or second order difference matrix. For example, for a p-spline with a first order random walk the penalty matrix is given by:



$$K = \begin{pmatrix} 1 & -1 & & & & \\ & -1 & 2 & -1 & & \\ & & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}$$

For full Bayesian inference, the unknown variance parameters  $\tau^2$  are also considered as random and estimated simultaneously with the unknown regression parameters. Therefore, hyperpriors are assigned to the variances  $\tau^2$  in a further stage of the hierarchy by highly dispersed (but proper) inverse Gamma priors  $p(\tau^2) \sim IG(a, b)$ .

$$p(\tau^2) = (\tau^2)^{-a-1} \exp\left(-\frac{b}{\tau^2}\right) \quad (5)$$

A common choices for the hyperparamters are small values, for example are  $a=$  and  $b=0.005$  (or  $b=0.0005$ ). Alternatively, one may take  $a=b=0.001$ (Fahrmeir, Kneib and Lang (2004)).

### Priors for Fixed Effects

In the absence of any prior knowledge for the parameter vector  $\gamma$  of fixed effects the study considered a diffuse prior  $\gamma_j \propto const, j=1...r$ . Another choice would be to work with a multivariate Gaussian distribution  $\gamma \sim N(\gamma_0, \Sigma)$ . In this study, diffuse priors was used for the fixed effects parameter  $\gamma$ .

### Bayesian P-spline

Any smoother depends heavily on the choice of smoothing parameters for p-spline in a mixed (fixed and continuous) framework. A closely related approach for continuous covariates is based

on the P-splines approach introduced by Eilers and Marx (1996). This approach assumes that an unknown smooth function  $f_j$  of a covariate  $X_j$  can be approximated by a polynomial spline of degree  $l$  defined on a set of equally spaced knots  $x_{\min} = \zeta_0 < \zeta_1 < \dots < \zeta_{d-1} < \zeta_d = x_{\max}$  within the domain of  $X_j$ . The domain from  $x_{\min}$  to  $x_{\max}$  can be divided into  $n'$  equal intervals by  $d'+1$  knots, Each interval will be covered by  $l+1$  P-splines of degree  $l$ , The total number of knots for construction of the P-spline will be  $d'+2l+1$ . The number of P-splines in the regression is  $d'+1$ . It is well known that such a spline can be written in terms of a linear combination of  $M_j = d + l$  P-spline basis functions  $B_m$ , i.e.,

$$f_j(x_{ij}) = \sum_{m=1}^{M_j} \beta_{jm} B_m(x_{ij}) \quad (6)$$

Here,  $\beta_j = (\beta_{j1}, \dots, \beta_{jm_j})$  corresponds to the vector of unknown regression coefficients. The  $n \times m_j$  design matrix  $\psi_j$  consists of the basic functions evaluated at the observations  $x_{ij}$ , i.e.,  $\psi_j(i, m) = B_m(x_{ij})$ . The crucial choice is the number of knots: for a small number of knots, the resulting spline may not be flexible enough to capture the variability of the data; for a large number of knots, estimated curves tend to over fit the data and, as a result, too rough functions are obtained. As a remedy, Eilers and Marx (1996) suggest a moderately large number of equally spaced knots (usually between 20 and 40) to ensure enough flexibility, and to define a roughness penalty based on first or second order differences of adjacent P-Spline coefficients to guarantee sufficient smoothness of the fitted curves. In our analysis, we will typically choose P-splines of degree 3 and 20 intervals, and second order random walk priors on

the P-splines regression coefficients. Hence, it is used to flexibly capture the variability of the data.

### **First and second order random walk priors**

Let us consider the case of a continuous covariate  $X$  with equally spaced observations  $x_i, i = 1, \dots, k, k < n$ . Suppose that  $x(1) < \dots < x(t) < \dots < x(k)$  defines the ordered sequence of distinct covariate values. Here  $n$  denotes the number of different observations for  $x$  in the data set. A common approach in dynamic or state space models is to estimate one parameter  $f(t)$  for each distinct  $x(t)$ ; *i.e.*, Define  $f(t) = f(x(t))$  and let  $f = (f(1), \dots, f(t), \dots, f(k))'$  denote the vector of function evaluations. Then a first order random walk prior for  $f$  is defined by:

$$f(t) = f(t-1) + u(t) \quad (7)$$

The second order random walk prior for  $f$  is defined by:

$$f(t) = 2f(t-1) - f(t-2) + u(t) \quad (8)$$

With Gaussian errors  $u(t) \sim N(0, \tau^2)$  and diffuse priors  $f(1) \propto \text{const}$ , or  $f(1)$  and  $f(2) \propto \text{const}$ , for initial values, respectively. A first order random walk penalizes abrupt jumps  $f(t) - f(t-1)$  between successive states and a second order random walk penalizes deviations from the linear trend  $2f(t-1) - f(t-2)$ . Random walk priors may be equivalently defined in a more symmetric form by specifying the conditional distributions of function evaluations  $f(t)$  given its left *and* right neighbors, e.g.  $f(t-1)$  and  $f(t+1)$  in the case of a first order random walk. Thus, random walk priors may be interpreted in terms of locally polynomial fits. A first order random walk

corresponds to a locally linear and a second order random walk to a locally quadratic fit to the nearest neighbors. Of course, higher order auto regressions are possible but practical experience shows that the differences in results are negligible Khaled (2010). The amount of smoothness is controlled by the additional variance parameter  $\tau^2$ , which corresponds to the smoothing parameter in a frequentist approach. The larger (smaller) the variances, the rougher (smoother) are the estimated functions. In addition, the variance  $\tau^2$  controls the degree of smoothness  $f$ .

$$(f_t | f_{t-1}, \tau^2) \propto N(f_{t-1}, \tau^2) \quad (9)$$

Additionally, random walk priors may be equivalently defined in a more symmetric form by specifying the conditional distributions of function  $f(t)$  given its left and right neighbors. That means, we can write the prior in (the first and second order random walk) in general form as

$$[f | \tau^2] \propto \exp\left(-\frac{1}{\tau^2} f' K f\right) \quad (10)$$

Here the design matrix  $K$  is the penalty matrix that penalizes too abrupt jumps between neighboring parameters. More often,  $K$  is not full rank and this implies that  $(f | \tau^2)$  follows a partially improper Gaussian prior

$$(f | \tau^2) \propto N(0, \tau^2 k^-)$$

Where,  $k^-$  is a generalized inverse of the penalty matrix  $K$ .

For the case of non-equally spaced observations random walk priors must be modified to account for non-equal distances  $\delta t = x(t) - x(t-1)$  between observations. Random walks of first order are now specified by  $f(t) = f(t-1) + u(t)$ ;  $u(t) \sim N(0; \delta t \tau^2)$ ; i. e. by adjusting error variances from  $\tau^2$  to  $\delta t \tau^2$ . Random walks of second order are defined by

$$f(t) = (1 + \frac{\delta_t}{\delta_{t-1}})f(t-1) - (\frac{\delta_t}{\delta_{t-1}})f(t-2) + u(t) \quad (11)$$

$u(t) \sim N(0, w_t \tau^2)$ , where  $w_t$  is an appropriate weight. Several possibilities are conceivable for the weights; see Fahrmeir and Lang (2001). However, in this analysis, we use a second order random walk prior for metrical covariates. Note that random walks are the special case of P-splines of degree zero.

### 2.3.3. Posterior Probability Distribution

Bayesian inference is based on the entire posterior distribution derived by multiplying the prior distribution  $\pi(\theta)$  of all parameters and the full likelihood function  $L(y|\theta)$ . For this case, let  $\theta$  be the vector of all unknown parameters, then the posterior distribution is given by:

$$\begin{aligned} \pi(\theta | y) &\propto L(y | \beta_1, \tau_1, \beta_2, \tau_2, \dots, \beta_p, \tau_p, \gamma) \prod_{j=1}^p \pi(\beta_j | \tau_j^2) p(\tau_j^2) \\ \pi(\theta | y) &\propto L(y | \beta_1, \tau_1, \beta_2, \tau_2, \dots, \beta_p, \tau_p, \gamma) \\ &\prod_{j=1}^p \frac{1}{(\tau^2)^{\text{rank}(K_j)/2}} \exp\left(-\frac{1}{2\tau^2} \beta_j' K_j \beta_j\right) \prod_{j=1}^p (\tau^2)^{-a_j-1} \exp\left(-\frac{b_j}{\tau_j^2}\right) \end{aligned} \quad (12)$$

In many practical situations (as is the case here) the posterior distribution is numerically intractable. To overcome this problem, Markov Chain Monte Carlo (MCMC) simulation technique is used to draw samples from the posterior. From these samples, quantities such as (posterior) mean, (posterior) standard deviation, and quantiles (which in turn, gives the associated credible interval) can be estimated. Bayesian inference via MCMC is based on updating full conditionals of single parameters or blocks of parameters, given the rest and the data. For the Gaussian response variable, the full conditionals for fixed effects and non-linear

effects are multivariate Gaussian. For the variance parameters, all full conditionals are inverse Gamma distribution. Straight forward calculations show that precision matrices for nonlinear effects are band matrices. In fully Bayesian inference, the unknown variance parameters  $\tau^2$  are considered as random and estimated simultaneously with the unknown regression parameters. Therefore, hyperpriors are assigned to the variances  $\tau^2$  in a further stage of the hierarchy by highly dispersed inverse Gamma priors  $p(\tau^2) \sim IG(a, b)$ .

### **3. RESULTS AND DISCUSSIONS**

#### **3.1. Descriptive Analysis**

The main purpose of this study was to determine statistically the correlates of child malnutrition in Ethiopia. The data were obtained from 2011 DHS (CSA, 2011). In this study a total of 7739 children under age five were considered for the analysis. Among these, 3350 (43.3%) were found to be stunted, with height-for-age Z-score less than -2.0. This shows that stunting prevalence is high in the country. The result displayed on Table 1 shows the percentages and counts of stunting status of children with respect to the categorical explanatory variables.

As one can see from Table 1, among the 7739 cases examined in this study 44.2 % of male children were stunted and 42.2 % of female children were stunted.

In this study, birth order is recorded into two categories: first to third birth and above. As one can see from the descriptive output in most of the household the number of children ever born is greater than three. In addition to this one can easily see from Table 1 as the birth order number increase it seems that children's nutritional status decrease.

As can be seen from Table 1, the highest prevalence of child malnutrition was observed among children whose preceding birth interval was less than 24 months (49.1%) unlike to the lowest

prevalence of child malnutrition which was recorded from children whose preceding birth interval is 48 and above (36.5 %).

In the rural areas, nearly half (45.4%) of all children are stunted. This figure is 30% for those who reside in urban areas.

From Table 1, nutritional status of a child varies by educational level of mother. The highest prevalence of malnutrition was observed from children whose mothers had no formal education (45.4%) as opposed to the lowest prevalence of child malnutrition which was recorded from children whose mothers have secondary and above educational level (27.1%). It seems that a mother with higher educational level had a child with better nutritional status.

Regarding husband/partner educational status, the highest probability of child to be stunted was observed for mothers whose husband/partner had no formal education (45.5%) and the lowest prevalence for child from a mother whose husband/partner had secondary and above educational level (28.8 %).

Table 1 also shows that the nutritional status of children varies by the households' economic status. The highest probability of stunting was observed among children from poor households (46.6 %) and the lowest was noticed for children cared by non-poor household. Households with more than or more children had the highest percentage of stunted children (44.6 %) unlike to households with less than five children (42.8 %).

Likewise, respondents were classified as those currently working and respondents who were not working. The survey report focuses on whether the mother was working at the time of the survey. Children of working mothers seemed less stunted (42 %).

The highest prevalence of child-stunting was observed in Amhara (52%), Affar (51.9%), Tigray (51.6%), and Benishangul-Gumuz (49.4 %) regions and it was the lowest in Addis Ababa and the Gambela region (26.5% and 28.2 %, respectively).

The evidence given in Table1 shows that the percentage nutritional statuses of children who had diarrhea recently seem to be high probability of stunting than those children which had no diarrhea recently. Likewise, one can observe that the percentages nutritional status of children who had fever recently seemed high probability of stunting than those children who have no fever recent (43.5%, 43.2 %, respectively).

The evidence given in Table 1 shows that the percentages nutritional status of children in households who had unprotected water seemed to be more severe of stunting than those children in households which used water from protected source.

**Table 1:** Distribution of Socioeconomic, Demographic, Health and Environmental related Characteristics vs Stunting (EDHS 2011).

Variable		Stunting Status (Z-Score)			
		Not Stunted		Stunted	
		Count	Percent	Count	Percent
Child Sex	Female	2204	57.8	62	42.2
	Male	285	55.7	738	44.3
Birth order of child	1-3	789	58.7	26	41.3
	4 and more	2600	55.4	2089	44.6
Preceding Birth Interval	<24 months	829	50.9	800	49.1
	24-47 months	2467	56.2	923	43.8
	48 and Above months	093	63.5	627	36.5
Wealth Index	Poor	225	53.4	852	46.6
	Medium or Rich	2264	60.2	498	39.8
Number of Household Members	1-5	560	57.2	67	42.8
	6-0	2645	56.5	2035	43.5



	and above	84	55.4	48	44.6
Mother's Education Level	No Formal Education	373	54.6	2637	45.4
	Primary Education	057	6.8	654	38.2
	Secondary and Higher	59	72.9	59	27.1
Husband/Partner's Education Level	No Formal Education	2358	54.5	967	45.5
	Primary Education	584	56.9	202	43.1
	Secondary and above	44	7.2	78	28.8
Current Work status of Mothers	Not work	3040	56.	2375	43.9
	Work	349	58.0	975	42.0
Residence	Rural	3653	54.6	3035	45.4
	Urban	736	70.0	35	30.0
Source of Water Supply	Unprotected	334	54.	2663	45.9
	Protected	255	64.6	687	35.4
Have Toilet Facility	No	2424	55.3	958	44.7
	Yes	965	58.5	392	41.5
Had Diarrhea Recently	No	3689	56.9	2797	43.1
	Yes	700	55.9	553	44.1
Had Fever in last two weeks	NO	3503	56.8	2669	43.2
	Yes	886	56.5	68	43.5
Region	Tigray	49	48.4	447	51.6
	Affar	348	48.	375	51.9
	Amhara	43	48.0	448	52.0
	Oromiya	735	58.8	54	41.2
	Somali	440	65.8	229	34.2
	Benishangul-Gumuz	346	50.6	338	49.4
	SNNP	626	55.3	505	44.7
	Gembela	394	7.8	55	28.2
	Harari	272	70.6	3	29.4
	Addis Ababa	22	73.5	44	26.5
Dire Dawa	279	6.2	77	38.8	

### 3.2. Results of Bayesian Semi-parametric Analysis

The whole analysis has been implemented using software BayesX (Belitz Andreas Brezger, 2009).

The fitted model was:

$$\begin{aligned} \text{Zscore} = & \hat{\beta}_0 + f_1(\text{MBI}) + f_2(\text{Mage}) + f_3(\text{Cage}) + \text{Csex}\gamma + \text{Res}\gamma_2 + \text{BORD}\gamma_3 + \text{pint}\gamma_4 + \text{Pbint2}\gamma_5 \\ & + \text{HHM}\gamma_6 + \text{HHM2}\gamma_7 + \text{Medu}\gamma_8 + \text{Medu2}\gamma_9 + \text{Pedu}\gamma_0 + \text{Pedu2}\gamma + \text{Region}\gamma_2 + \text{tfacility}\gamma_3 \\ & + \text{Drrh}\gamma_4 + \text{fevery}\gamma_5 + \text{windex}\gamma_6 \end{aligned}$$

#### 3.2.1. Linear Fixed Effects

Table 2 gives results for the fixed effects (categorical covariates) on the nutritional status of children under age five in Ethiopia. The output gives posterior means, posterior median along with their standard deviations and 90% credible intervals. Since the 90% confidence interval do not include zero, Sex of child, birth order of child, birth interval, place of residence, region, mothers education level, toilet facility, household members, household economic status, husband/partners educational level, diarrhea status of child and fever status of child were found statistically significant at 5% significance level. But, source of drinking water and respondent current work status were found statistically insignificant.

From Table 2, one can observe that having an educated mother (at least primary education) contributes to better nourishment for children under five age which has also been found in other studies (Mohammed, 2008; Khaled, 2007). In this study, the relative chance to be stunted for the children was found to decrease with the increase of mother's educational level. The children of illiterate mothers and those with incomplete primary education were more likely to be malnourished as compared to mothers with secondary education and higher. The findings

support the statement that educated mothers were more conscious about their children's health Mohammed (2008). Literate mothers can easily introduce new feeding practices scientifically, which helps to improve the nutritional status of children. And also the analysis shows that female children are better nourished than male children.

On the other hand, one can observe that stunting is higher for children of higher birth order (other than first born), larger household's members, and for child residents in the rural areas. Also we observe that larger household is not conducive for better nourishment of children. However, one may interpret that while larger households provide more care to children (mostly by elder members of the household in a joint or extended family setting), there seemed to be a simultaneous competition for resources within the same larger household size. This competition for limited resources may be responsible for worsening of nutritional status for the children of a larger household size. Mother's current working status has no effect on child malnutrition. This study suggests that child malnutrition may not be influenced by current working status of mothers since chronic malnutrition (stunting) is long term deficiency of nutrients.

Household factors are strong indicators of children's stunting status. Children in households having higher income have better nutritional status than that of lower income households. The results indicate that the risk of children being stunted decreased with the increase of household wealth index. The children of households having the lowest wealth index were more likely to be stunted than those of household with the highest wealth index. Working status of mothers which might increase economic status of household had insignificant effect on child nutritional status.

The prior birth interval matters for the nutritional status of the child. The analysis showed that children born after a long birth interval were better off than other children. This statement is in

agreement with Das et. al. (2008) which found that, children with a birth interval of 2 or more years were less likely to be stunted. This may be due to the fact that the parents can take better care of fewer numbers of children and could provide adequate breast milk due to recovery of nutritional status between births. The results also indicate that source of water had no significant effect on child stunting status.

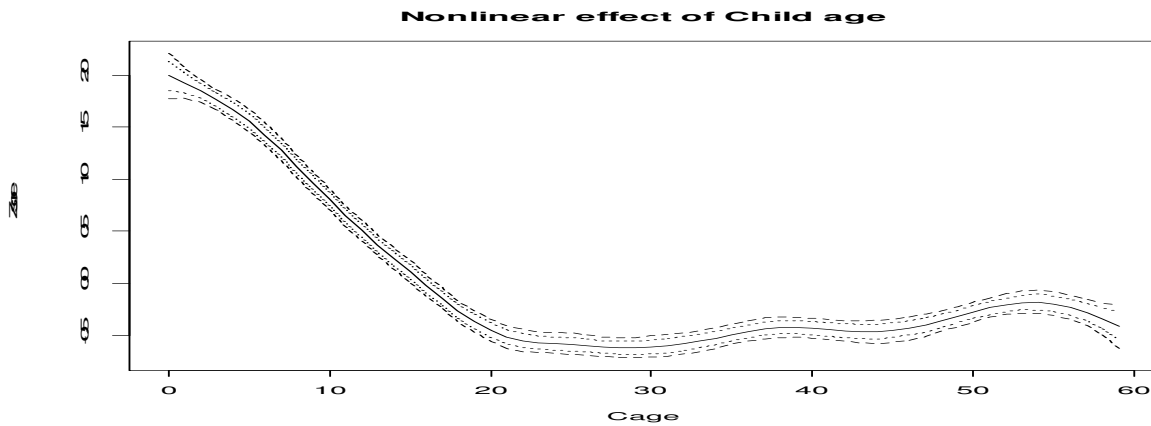
**Table 2:** Posterior Results of the Categorical Covariates.

Variable	Pmean	PStd. Dev.	10% quant.	pMedian	90% quant.
Const	-1.68*	0.10	-1.81	-1.67	-1.55
Csex	-0.085*	0.034	-0.15	-0.08	-0.04
BORD	-0.12*	0.05	-0.18	-0.12	-0.05
Residence	0.19*	0.07	0.11	0.19	0.28
Medu0 (ref)	-	-	-	-	-
Medu1	0.13*	0.04	0.07	0.13	0.19
Medu2	0.3*	0.13	0.15	0.32	0.48
Swat	0.05	0.05	-0.01	0.05	0.11
Tfacility	-0.07*	0.04	-0.12	-0.07	-0.02
HHM 0(ref)	-	-	-	-	-
HHM1	0.22*	0.045	0.17	0.22	0.28
HHM2	0.39*	0.09	0.27	0.39	0.51
Windex	0.11*	0.04	0.05	0.11	0.16
Pedu0(ref)	-	-	-	-	-
Pedu1	0.06*	0.04	0.013	0.064	0.12
Pedu2	0.33 *	0.08	0.23	0.33	0.43
Mwork	-0.006	0.04	-0.05	-0.003	0.04
DRRH	-0.14*	0.05	-0.2	-0.14	-0.07
Fever	-0.011 *	0.05	-0.16	-0.11	-0.05
Pbint(ref)	-	-	-	-	-
Pbint1	0.06*	0.04	0.005	0.06	0.12
Pbint2	0.2*	0.06	0.15	0.2	0.33
Region	0.04 *	0.006	0.03	0.04	0.05

### **3.3. Non-linear Effects under Generalized Additive Linear Regression Models**

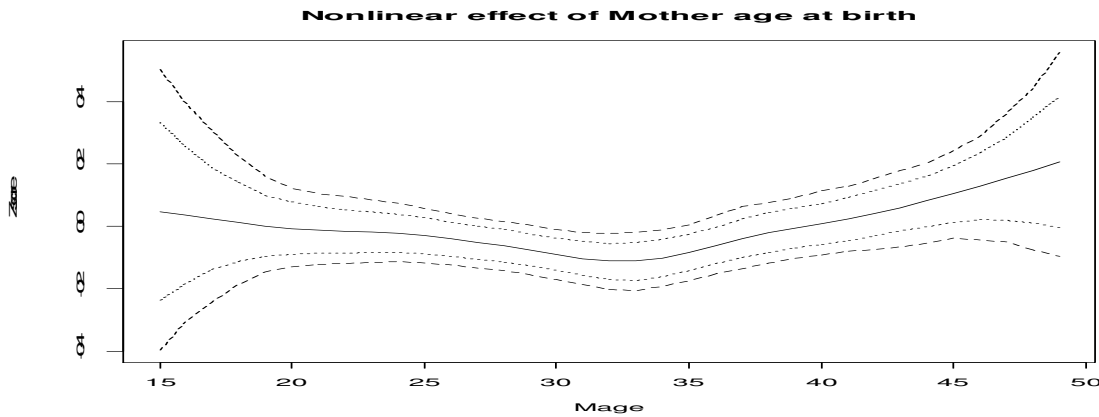
Nonlinear effects represented by smoothed functions, are commonly interpreted graphically. Figure 1 shows the smooth function of the children age versus height-for-age z-score. The posterior means together with 80% and 95% point wise credible intervals are shown. One can observe that the influence of a child's age on its nutritional status is considerably high in the age range between the ages of 0-20 months with decreasing trend; and then stabilizes.

As suggested by the nutritional literature, one can able to distinguish the continuous worsening of the nutritional status up until about 20 months of age. This deterioration set in right after birth and continues, more or less linearly, until 20 months. After 20 months the effect of age on stunting stabilizes at a low level. Through reduced growth and the waning impact of infections, children were apparently able to reach a low-level equilibrium that allows their nutritional status to stabilize (Khalid, 2007; Mohammed, 2008; Kandala, 2010; Belitz et.al, 2007, Khalid, 2010).



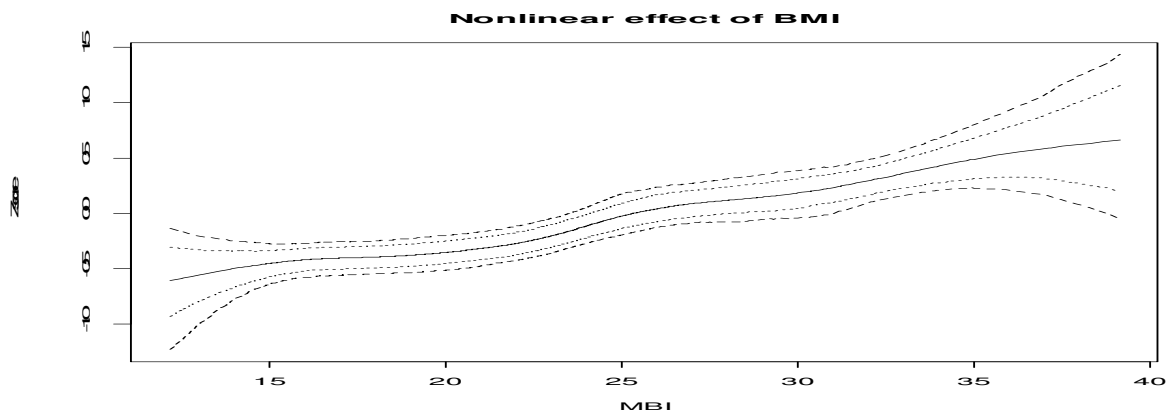
**Figure 1:** Non-linear Effects of Child’s Age on Nutritional Status of a Child.

Figure 2 displays nonlinear effects of mother’s age at birth. It shows the posterior means together with 80 % and 95% point-wise credible intervals. It is evident from the analysis that increasing age of mother at birth reduces stunting status of children. That is younger mothers tend to have more stunted children than older mothers. Mother age at birth shows significant effect on stunting status of children under age of five years old. The effect of mother’s age on her child’s stunting status (other constant) negatively increase as her age increase up to 33 years and then after the effect of mothers age on her child’s stunting status positively increase.



**Figure 2:** Non-linear Effects of Mother’s Age on Nutritional Status of a Child.

Figure 3 shows the flexible modelling of the effect of the BMI of the mother versus stunting. The posterior means together with 80 % and 95% point wise credible intervals are displayed. Bearing in mind at the mother’s BMI and its impact on the level of nutritional status, it was found that the influence had a regular pattern. The stunting status of a child improves as mothers BMI increases. In general, the figure shows that BMI had a significant effect on child nutritional status.



**Figure 3:** Non-linear Effects of Mother’s Body Mass Index on Nutritional Status of a Child.

### 3.4. Discussion

This study was intended to identify the determinants of the stunting status of children under five years old in Ethiopia based on EDHS 2011 data. The nutritional status was measured by the height-for-age or stunting. Accordingly, Bayesian semi-parametric regression analysis on stunting was employed to identify flexibly the effect of covariates on nutritional status of the children. In this study, the chronic malnutrition status (stunting) was analyzed based on the modified anthropometric measurement indicators of the nutritional status of children calculated

using new growth standards published by the World Health Organization in 2006. The results obtained are discussed as follows. The total number of children covered in the present study was 7739, among which 43.3 % were stunted. The Bayesian semi-parametric analysis revealed that the covariates: sex of child, birth order of child, place of residence, region, previous birth interval, mothers education level, toilet facility, household members, household economic status, husband/partners educational level, diarrhea and fever were found statistically significant. But, source of drinking water and mother's current work status were found statistically insignificant (though not expected).

Preceding birth interval is an important demographic variable that affects nutritional status of children. As the preceding birth interval increases, the nutritional status of a child increases. This finding is confirmed by most of previous studies (Kandala, 2009; Khaled, 2010; Mohammed, 2008). The significant and higher risk of stunting among children of lower preceding birth interval could be due to uninterrupted pregnancy and breastfeeding, since this drains women's nutritional resources. Close-spacing may also have a health effect on the previous child who may be prematurely weaned if the mother becomes pregnant too early again.

The fact that mother education level was significant is in agreement with different reports undertaken on the same problem. According to Sasha (2009), with an increase in mothers' educational level, incidence of malnutrition among young children decreases. And this factor is then directly associated to the women's own nutritional status and quality of care they receive (Smith et. al, 2003). The educated mothers are more conscious about their children's health. Literate mothers can easily introduce new feeding practices scientifically, which help to improve their children



nutritional status (Das *et. al.*, 2008). Greater formal education levels achieved by both mothers and fathers were associated with decreased odds of child stunting (Semba *et. al.*, 2008).

The study suggests that insufficient food intake may be not affected by the current working status of mothers. The results are consistent with some previous studies and not consistent with others. Some studies reported that when mothers are working, the household income is increased and the access to better food will be increased, as well as the access to a quality level of medical care. On the other hand, when mothers stay outside the home, it curtails the duration of full breastfeeding and necessitates supplementary feeding, usually by illiterate care-takers, which might affect the health of children negatively (Khaled, 2010).

Household economic status is also an important socio-economic variable that affects nutritional status of children in Ethiopia. Children in poor households were found to be at a higher risk of malnutrition problem than children from rich households. This finding is consistent with other studies (Smith *et. al.*, 2005; Woldemariam & Timotewos, 2002). The study indicated that better off households had better access to food and higher cash incomes than poor households, allowing them a quality diet, better access to medical care and more money to spend on essential non-food items such as schooling, clothing and hygiene products.

As the study has revealed that, household size is an important variable that affects nutritional status of children. The prevalence of stunting increased with increasing household size. Households with more than and above child had a higher percentage of stunted children (44.6%) compared with households with less than five children (42.8 %). Large household size is not conducive for better nourishment of children. However, we may interpret that while larger households provide more care to children (mostly by elder members of the household in a joint

or extended family setting), there seemed to be a simultaneous competition for resources within the same larger household size. This competition for limited resources may be responsible for worsening of nutritional status for the children of a larger household size. This result is in agreement with a study under taken by Rahman & Chowdhury (2007) which stated that children living in a household with only one child have a lower risk of stunting than children who live in households with more than one child. The total number of children within a household influences the resources available to each child, in terms of financial, time and attention. In a crowded household, exposure of an individual child to infection is also increased (Sereebutra *et al.*, 2006).

As shown in the analysis, urban children were less likely to be malnourished than their rural counterparts because the quality of health environment and sanitation is better in urban areas, whereas, the living condition in rural areas were associated with poor health condition, and lack of personal hygiene, which were the risk factors in determining malnutrition. This is consistent with some studies, where mothers' place of residence has a statistical significant effect on children nutritional status (Kandala et al, 2006; Woldemariam & Timotiows, 2002).

The findings of this study also showed that children who had diarrhea two weeks before date of survey are vulnerable to malnutrition problem than those who had not. This finding is consistent with other studies (Khaled, 2007; WHO, 2011; Birhan, 2010). This may be due to the fact that diarrhea accelerates the onset of malnutrition by reducing food intake and increasing catabolic reactions in the organism. Diarrhea also affects both dietary intake and utilization, which may have a negative effect in child nutritional status. The type of toilet used by a household is an indicator of household wealth and a determinant of environmental sanitation. This means that

poor households were less likely to have sanitary toilet facilities. In consequence, these results increased risk of childhood diseases, which contribute to malnutrition.

It is evident that increasing age of mother's at birth reduces stunting and shown that children whose mothers are older age group were better in their stunting status as compare to children whose mothers are in the younger age group. The mother's age at birth of the index child was found to be a statistically significant predictor of children's stunting status in a number of previous studies, where it was found that children born to mothers between the ages of 20 and 29 years were more likely to have children that suffered from negative nutritional outcomes than those children born to older mothers (Som et al 2006). Children whose mothers are older than 30 years of age are better in their stunting status as compare to children whose mothers are in the younger age group (Mohammed, 2008).

In the study, the effect of the age of the child is obviously nonlinear and decreasing between birth and an age of about 20 months and then stabilizes. That means the stunting of children increases until 20 months. This continuous worsening of the nutritional status may be caused by the fact that most of the children obtain liquids other than breast milk already shortly after birth. After 20 months a relatively stable, low level is reached. However, it reaches its minimum level between ages 20-30 months, then rises again and stabilizes thereafter at a middle level with a bump till 5 years. Previous studies also confirm this (Kandala, et. al. 2007; Khalid, 2006; Mohammed, 2008; Kandala, 2011).

Mother's body mass index is defined as her weight in kilo-grams divided by her square of her height in meters. Mothers with low BMI value are themselves malnourished and are therefore likely to have undernourished children. The same finding is also found in a number of studies. Mothers with low BMI on average give birth to babies of low birth weight (Khalid, 2007; Mohammed, 2008).

#### 4. CONCLUSIONS

The main objective of this study was to identify the most important predictors of under five years old children stunting status in Ethiopia using Bayesian Semiparametric regression model. The study revealed that socio-economic, demographic and health and environmental variables have significant effect on the stunting status of children in Ethiopia. Using Bayesian Semiparametric regression model, the predictors, sex of child, birth order of child, preceding birth interval, place of residence, region, mother's education level, toilet facility, household members, household economic status, husband/partners educational level, diarrhea and fever are the most important determinants of child stunting status in the country.

The study showed that children from uneducated mother, uneducated partner of mother, lower preceding birth interval (less than 24 months), higher birth order, economically poor household, large household size are more vulnerable to stunting problem in Ethiopia. The findings of this study also showed that children who had diarrhea and fever for two weeks before the date of survey are significantly vulnerable to stunting problem than those who had not. Male children are more vulnerable to stunting problem than female children.

The study analyzed how the continuous covariates: child's age, mother's age at birth and mother's BMI affect stunting status of the child. The stunting status of a child improves as mother's body mass index increases. And children are at high risk of stunting problem during the first 0-20 months of their life and then stabilize moderately with bump. The study also revealed that children whose mothers are in the older age group are better in their nutritional status as compare to children whose mothers are in the younger age group.

## 5. REFERENCES

- Belitz, C.; Brezger A., Kneib T.; Stefan L. (2009). BayesX Software for Bayesian Inference in Structured Additive Regression. Department of Statistics, Ludwig Maximilians University Munich. Version, 2.0..
- Birhan, F. (2010). Determinants of Nutrition and Health Status of Children in Ethiopia: A Multivariate Multilevel Linear Regression Analysis. Addis Ababa University.
- Brezger, A. and Lang, S. (2006). Generalized Structured Additive Regression based on Bayesian P-Splines. *Computational Statistics and Data Analysis*, Vol 50,P. 967-99.
- Das, S., Hossain; M.Z., & Islam, M.A. (2008). Predictors of Child Chronic Malnutrition in Bangladesh. *Proc.Pakistan Acad. Sci.* 45(3): P.37-55.
- Central Statistical Agency (2011). Ethiopian Demographic and Health Survey. Addis Ababa
- Fahrmeir, L. and Lang, S. (2001). Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. *Applied Statistics (JRSS C)*, Vol 50, P. 20-220.
- Fahrmeir, L. and Lang, S. (2000). Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. To appear in *Ann. Inst. Statist. Math.*
- Hastie, T. and Tibshirani, R. (2000). Bayesian Back Fitting. (To Appear in *Statistical Science*).
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Kandala, N.B. (2011). Malnutrition Among Children under the Age of Five in the Democratic Republic of Congo (DRC): Does Geographic Location Matter?
- Kandala, N.B; Fahrmeir L; Klasen S; Priebe J (2009). Geo-additive Models of Childhood Undernutrition in Three Sub-Saharan African Countries. *Population, Space and Place*.

- Kandala ,N.B; S. Lang; S. Klasen, and L. Fahrmeir (2006). Semiparametric Analysis of the Socio-Demographic Determinants of Undernutrition in Two African Countries. Research in Official Statistics, EUROSTAT, Vol. 4 No.:P. 8-00.
- Kandala, N.B.; Fahrmeir, L and Klasen, S. (2002). Geo-additive Models of Childhood Undernutrition in Three Sub-Saharan African Countries. Sonderforschungsbereich 386: Discussion Paper no. 287. Deutsche Forschungsgemeinschaft.
- Khaled, K. (2010). Child Malnutrition in Egypt Using Geoadditive Gaussian and Latent Variable Models.
- Kibel ,M.; Saloojee; H. & Westwood, T. (2007). Child Health for All. (4th Edition). Oxford University Press: Cape Town (South Africa).
- Klasen, S. and Moradi, A. (2000). The Nutritional Status of Elites in India, Kenya and Zambia: An Appropriate Guide for Developing Reference Standards for Undernutrition? Sonderforschungsbereich 386: Discussion Paper no. 27. Deutsche Forschungsge-  
meinschaft.
- Maleta, K. (2006). Epidemiology of Undernutrition in Malawi, Chapter 8 in The Epidemiology of Malawi, Edited by Eveline Geubbles and Cameron Bowie.
- Mila, A.L.; Yang, X.B. and Carriquiry, A.L. (2003). Bayesian Logistic Regression of Clinical Epidemiology for Uncertainty in Parameter Estimation. Basic Science for Clinical Medicine: Little, Brown and Company, Boston.
- Mohammad, A. (2008). Gender Differentials in Mortality and Undernutrition in Pakistan: Peshawar (Pakistan).
- National Nutrition Strategy (2009). Ethiopia National Nutrition Strategy Review and Analysis of Progress and Gaps: One Year On May 2009.
- Rahman, A. & Chowdhury, S. (2007). Determinants of Chronic Malnutrition Among Preschool Children in Bangladesh. *Journal of Biosocial Science* 39(2):P.6-73.

- Sasha, F. (2009). An Analysis of Under-Five Nutritional Status in Lesotho: The Role of Parity Order and Other Socio-Demographic Characteristics.
- Semba, R. D.; de Pee, S., Sun, K.; Sari, M.; Akhter, N., & Bloem, M.W. (2008). Effect of Parental Formal Education on Risk of Child Stunting in Indonesia and Bangladesh: A Cross Sectional Study. *Lancet* 37(9609):P.322-328.
- Sereebutra, P.; Solomons, N.; Aliyu, M.H., & Jolly, P.E. (2006). Socio-Demographic and Environmental Predictors of Childhood Stunting in Rural Guatemala. *Nutrition Research*.
- Smith, L. C. and Haddad L. (1999). Explaining Child Malnutrition in Developing countries: a Cross Country Analysis. International Food Policy Research Institute, FCND Discussion paper, USA.
- Smith, L. C. and Haddad, L. (2000). Overcoming Child Malnutrition in Developing Countries: Past Achievements and Future Choices. International Food Policy Research Institute, Discussion Paper 30, USA.
- Som, S.; Pal M.; Bhattacharya; B. Bharati, S. & Bharati, P. (2006). Socio-Economic Differentials in Nutritional Status of Children in the States of West Bengal and Assam, India. *Journal of Biosocial Sciences. Vol.38: P.625-642*.
- UNICEF (2009a). Tracking Progress on Child and Maternal Nutrition, a Survival and Development Priority.
- UNICEF (2009b). To Reduce the Rate of Low Birth Weight by at Least One Third. *Child Info, 2009; The State of the World's Children. Maternal and Newborn Health*.
- Woldemariam, G . and Timotiows, G. (2002). Determinants of Nutritional Status of Women and Children in Ethiopia
- WHO (2011). World Health Statistics, (The Millennium Development Goals Report), United Nations 2011.
- WHO (1995). Physical Status: The Use and Interpretation of Anthropometry. WHO Technical Report Series No. 854. Geneva.
- World Bank (2007). Nutritional Failure in Ecuador: Causes, Consequences, and Solutions. The World Bank: Washington, DC.