ARC Journal of
Academic Research
www.africaresearchcorps.com
Volume 1, Issue 1, pp. 01-10

Bayesian Regression Method with Gaussian and Binomial Links for the Analysis of Nigerian Children Nutritional Status (Stunting)

[*1]Lasisi, K.E.

[2]Nwaosu, S.C.

[1]Abdulhamid, B.M.

*[Corresponding author] E-mail: klasis2000@yahoo.com | Tel: +2348038883956, +2348023826362
[1]Mathematical Science Programme, Abubakar Tafawa Balewa University, P.M.B. 0248, Bauchi, Bauchi State, Nigeria
[2]Department of Mathematics/Statistics, University of Agriculture, Makurdi

ARC Journal of Academic Research (ARC-JAR)
Volume 1, Issue 1, (June, 2015) pages 01 - 10
www.africaresearchcorps.com

Bayesian Regression Method with Gaussian and Binomial Links for the Analysis of Nigerian Children Nutritional Status (Stunting)

*[1]Lasisi, K.E., [2]Nwaosu, S.C. and [1]Abdulhamid, B.M.
[1]Mathematical Science Programme, Abubakar Tafawa Balewa University, P.M.B. 0248, Bauchi, Bauchi State, Nigeria
Department of Mathematics/Statistics, University of Agriculture, Makurdi

Abstract: *Children's nutritional status is a reflection of their overall health. Malnutrition is associated with more than half of all children deaths worldwide. A study into geographical variability of nutritional status of children in Nigeria was observed from geostatistical mapping and a continuous covariates stunting (height for age) that exhibit pronounced non-linear relationships with the response variable was analysed. To properly account for stunting effects on child's age, sex, their place of resident, mothers' educational levels, parents' wealth index, regions and state of the child, kriging and additive models were merged using modified Cox model. The resulting Generalized Additive Mixed Model (GAMM) representation for the geoadditive model allows for fitting and analysis using BayesX software. The Multiple Indicator Cluster Survey 3 (MICS3) data set contains several variables. Only those that are believed to be related to nutritional status were selected. All categorical covariates are effect coded. The child's age is assumed to be nonlinear; the state is spatial effect while other variables are parametric in nature.*

Key words: *Binomial, Bayesian, Gaussian, Stunting, and Geostatistical*

Corresponding author: **E-mail:** klasis2000@yahoo.com | **Tel:** +2348038883956, +2348023826362

1. Background of the Study

Child mortality reflects a country's socio-economic development and quality of life. In developing countries, mortality rates are not only influenced by socio-economic, demographic and health variables but they also vary considerably across regions. Worldwide, an estimated 852million people are undernourished with about 815millions living in developing countries (FAO, 2004). The Millennium Development Goal is to reduce by half the proportion of people who suffer from hunger between 1990 and 2015. The World Fit for Children goal is to reduce the prevalence of malnutrition among children under five years of age by at least one-third (between 2000 and 2010), with special attention to children under 2 years of age. A reduction in the prevalence of malnutrition will assist in the goal to reduce child mortality (SOWC, 2007).

Geostatistics is concerned with the problem of producing a map of quantity of interest over a particular geographical region based on, usually noisy, measurements taken at set of locations in the region. Classical parametric regression models for analyzing child mortality or survival have severe problems with estimating small area effects and simultaneously adjusting for the covariates, in particular when some of the covariates are nonlinear and time-varying. Usually a very high number of parameters will be needed for modeling purposes, resulting in rather unstable estimates with high variance. Therefore, flexible semi-parametric approaches are needed which allow one to incorporate small area spatial effects, nonlinear and time-varying effects of covariates and usual linear effects in a joint model.

The Cox proportional hazards model is a commonly used method when analyzing the impact of covariates on continuous survival times. In its classical form, the Cox model was introduced in the setting of right-censored observations. However, in practice other sampling schemes are frequently encountered and therefore extensions allowing for interval and left censoring or left truncation are clearly desired. Furthermore, many applications require a more flexible modeling of covariate information than the usual linear predictor. Further extensions should allow for time-varying effects of covariates or covariates that are themselves time-varying. Such models relax the assumption of proportional hazards.

1.1. Nigeria Children Nutritional Status

In Nigeria, one in four that is, 25 percent of children under five are moderately or severely underweight and eight percent are classified as severely underweight. More than a third (34 percent) of children are moderately or severely stunted or too short for their age and 11 percent are moderately or severely wasted or too thin for their height. Severely stunted and severely wasted children are 19 and three percent respectively. The figure shows some curious trends in children malnutrition figures in Nigeria over the years (SOWC, 2007)**.**

From MICS surveys of 1999 and 2007 some decline in moderate underweight and in moderate wasting prevalence respectively between 1999 and 2007 is contrasted by corresponding increase in moderate stunting prevalence over the same period. Similar pattern is observed in respect of severe malnourishment figures. It also shows that severe underweight prevalence among children in Nigeria in 2007 is 8 percent and that prevalence rate depends on residence, age of the child, education of the mother, wealth status of the household and geopolitical zone; it declines from the poorest to the richest wealth index quintile of households, from the poorly educated to the highly educated and from the rural to the urban residents. The severe underweight prevalence is also observed to be higher in the northern zones than the south. (SOWC, 2007)

2. Methodology

The analysis was carried out using BayesX software package, which permits Bayesian inference based on MCMC simulation techniques. The statistical significance of apparent associations between potential risk factors and the stunting components was used to evaluate the significance of the posterior mean determined for the fixed effects or the categorical data, while non-linear effects and spatial effects were analysed using the estimation of spatial effects based on Markov random fields, stationary Gaussian random fields, and two-dimensional extensions of penalized splines properties of the programme and viewing the map through GSview 4.9 software. We also run a sensitivity analysis for the choice of priors. Standard choices for the hyper-parameters are a

= b = 0:001, with 25000 iteration and burn-in period of 5000, there are 17093 observations.

The Z-score: A z-score is a measure of how far a child's WHZ, WAZ and HAZ measurements are compared to an international reference value for a normal population.That is by comparing the body measurements of a given child with those of healthy children of the same height or age, we can classify his or her nutritional status. The healthy groups of children that are used for comparison are known as the **reference population**.A child below a certain weight for a specified height, or age would be considered malnourished. NCHS/WHO/CDC reference table, (WHO,2010) has been used as the international reference standard.Based on the anthropometric index of height-for-age of children under five (years) measured as z-scores (that is the standard deviations from the median of the reference population) such that

$$z_i = \frac{HAi - Med}{s.d}$$

Where HA_i is the anthropometric index of height-for-age for a child i, Med and s.d. are the median and the standard deviation of the reference population respectively.

3. The Models

Since the publication of the seminal paper of Cox (1972) cited in (Kneib and Fahrmeir, 2004) influences of covariates on survival times are commonly described by a regression model for the hazard rate. The Cox proportional hazards model assumes the multiplicative structure

$$\lambda(t,v) = \lambda_0(t) \exp(v'\gamma), \tag{3.1}$$

where$\lambda_0(t)$ is an unspecified smooth baseline hazard rate and $v'\gamma$ is a linear predictor form of covariates v and regression coefficientsγ. On the line of additive regression models, the Cox model can be extended to

$$\lambda_i(t) = \exp(\eta_i(t)), \qquad i = 1, \ldots, n, \tag{3.2}$$

where$i$ is an observation index and $\eta_i(t)$ is a geoadditive predictor of the form

$$\eta_i(t) = v_i'\gamma + g_0(t) + \sum_{l=1}^{L} g_l(t)u_{il} + \sum_{j=1}^{J} f_j(x_{ij}) + f_{spat}(S_i) \tag{3.3}$$

Here $g_0(t)$ = $\log(\lambda_0(t))$ is the log-baseline hazard, $g_l(t)$ represents time-varying effects of covariates u_{il}, $f_j(x_{ij})$ are nonlinear effects of continuous covariates, $f_{spat}(s_i)$ is a spatial effect, and $v_i'\gamma$ corresponds to covariate effects that are modeled in the usual parametric way. Nonparametric effects $f_j$as well as time-varying effects $g_0(t)$ and $g_l(t)$ are estimated based on penalized splines.

3.1 Gaussian Processes

It has been shown that many Bayesian regression models based on neural networks converge to Gaussian processes in the limit of an infinite network (Neal 1996). This has motivated examination of Gaussian process models for the high-dimensional applications to which neural networks are typically applied (Williams and Rasmussen 1996). The empirical work of Rasmussen (1996) has demonstrated that Gaussian process models have better predictive performance than several other nonparametric regression methods over a range of tasks with varying characteristics. The conceptual simplicity, flexibility, and good performance of Gaussian process models make them very attractive for a wide range of problems. Hence, the process was modified to fit into the Generalized Additive Mixed Model (GAMM) of Bayesian method. Furthermore, the response variables of interest are defined for Gaussian process as:

$$y \sim N(\mu, \Sigma), \text{ and } y \sim f(\gamma),$$
$$\text{where } \gamma = \beta_o + \beta_i X_i + \ldots + \beta_k X_k + f(Z) \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Where yis the regression response for stunting with respect to Gaussian regressions. Andγ is the geoadditive predictor which can be specified for a particular child *i*. The β_0, $\beta_i X_i$ and f(Z)represent the estimates of the unknown nonlinear smoothing effects of the metrical covariates child's age (cage), a vector of the fixed effect parameters and the spatial effect respectively. To enhance identifiability, functions are centred about zero, thus the fixed effect parameters automatically include an intercept term γ_0.Stunting = HAZ (Normal regression), where: HAZ - Height for Age Z-score.

3.2 Binomial regression

Bayesian approaches to estimating binary regression models took a sizable step forward with Zellner and Rossi (1984). They examined the generalized linear models (GLMs) $h[E(y_i)] = x_i'\beta$, where $\{y_i\}$ are independent binary random variables, x_i is a vector of covariates for y_i, and h(.) is a link function such as the probit or logit. They derived approximate posterior densities both for an improper uniform prior on β and for a general class of informative priors, giving particular attention to the multivariate normal. Daniels and Gatsonis (1999) used such modeling to analyze geographic and temporal trends with clustered longitudinal binary data. Biggeri*et al* (2004) used it to investigate the joint contribution of individual and aggregate (population-based) socioeconomic factors to mortality in Florence. They illustrated how an individual-level analysis that ignored the multilevel structure could produce biased results.

Hence, the need to consider the multilevel analysis as against the individual level analysis used with the Gaussian process, therefore the Binary regression was modified to fit into the Generalized Additive Mixed Model (GAMM) of Bayesian method. Furthermore, the response variables of interest are defined for Binomial process as:

$$y \sim B(n, p), \text{ and } y \sim f(\eta),$$
$$\text{where } \eta = f(cage_i) + f_{spat}(s_i) + v_i'\gamma$$

Where yis the regression response for stunting with respect to Binomial regressions. And where $\eta$is the geoadditive predictor which can be specified for a particular child *i*. The $f(cage_i)$, $f_{spat}(s_i)$ and γrepresent the estimates of the unknown nonlinear smoothing effects of the metrical covariates cage(child's age), the spatial effect and a vector of the fixed effect parameters. To enhance identifiability, functions are centred about zero, thus the fixed effect parameters automatically include an intercept term γ_0.

$$\text{Stunting} \begin{cases} 1 & \text{if HAZ} < -2 \\ 0 & \text{otherwise} \end{cases}$$

4. Results and Discussion

Nigerian children nutritional data was analyzed with the aim of assessing the influence of some covariates on the response variable (malnutrition). Since the Multiple Indicator Cluster Survey3 (MICS3) data set contains several variables, only those that are believed to be related to nutritional status were selected. All categorical covariates are effect coded. The child's age is assumed to be nonlinear; the state is special effect while other variables are parametric in nature.

The variables are defined as follows:

Area: Sector

Rural (reference location)

Urban

Geopolitical Zones

NC – North Central (reference zone)

NE – North East

NW – North West

SE – South East

SS - South South

SW – South West

Mother's Education

None (reference educational level)

Primary

Secondary

Non-Standard Curriculum (non-std)

Parents' Wealth Index Quintiles

Poorest (reference wealth index)

Windex2 – Second rich

Windex3 – Middle rich

Windex4 – Fourth rich

Windex5 – Richest

Sex:

Female (reference sex)

Male

4.1.1 Stunting Gaussian Regression

>f.regress stunting =state_rec(spatial, map=m, lambda=0.1) + CAGE(psplinerw2) + urban + WIndex2 + WIndex3 + WIndex4 + WIndex5 + primary + secondary + non_stdcur + UF11 + male + NEast + NWest + SEast + SSouth + SWest, iterations=25000 burnin=5000 step=20 family=gaussian predict using d

4.2.1 Stunting Binomial Regression

>f.regressstuntbin = state_rec(spatial, map=m, lambda=0.1) + CAGE(psplinerw2) + urban + WIndex2 + WIndex3 + WIndex4 + WIndex5 + primary + secondary + non_stdcur + UF11 + male + NEast + NWest + SEast + SSouth + SWest, iterations=25000 burnin=5000 step=20 family=binomial predict using d

Table 1- Stunting: Gaussian and Binomial Regression Analysis

Variable	**GaussianStunting**Odds ratio	95% Confidence Interval		**BinomialStunting**Odds ratio	95% Confidence Interval	
		lower limit Odds	upper limit Odds		lower limit Odds	upper limit Odds
Urban	0.7313	0.6507	0.8231	1.1649	0.2507	0.9289
Wealth Index2	1.0145	0.8902	1.1557	0.9582	1.0277	1.3277
Wealth Index3	1.0090	0.8845	1.1684	0.9251	0.8488	1.0843
Wealth Index4	1.0891	0.9241	1.2761	0.8877	0.8060	1.0542
Wealth Index5	1.4260	1.1748	1.7209	0.7994	0.7615	1.0475
Primary	0.9967	0.8837	1.1181	1.1058	0.6563	0.9801
Secondary	1.1416	0.9977	1.3041	0.9276	0.9824	1.2436
Non-std. curriculum	1.0194	0.7421	1.4111	0.8942	0.8046	1.0578
Male	0.8650	0.8018	0.9451	1.1086	0.6721	1.1862
Northeast	0.6835	0.5100	0.9094	1.2018	1.0211	1.2049
Northwest	0.2414	0.1441	0.3818	1.5622	0.8518	1.6277
Southeast	1.2425	0.7142	2.2780	0.6944	0.8092	2.9716
Southsouth	1.0883	0.6385	1.9342	0.5603	0.3141	1.7379
Southwest	0.9666	0.5416	1.7729	0.7288	0.2386	1.2764

The above tableshows that at 95% Confidence Interval, the prevalence of stunting (Gaussian) was higher among children living in the rural area with 27%more, while severe stunting (Binomial) was about 16.5% higher in children living in urban area. When comparing the two situations, we discovered that stunting which is a reflection of chronic malnutrition as a result of failure to receiveadequate nutrition over a long period and recurrent or chronicillness is prevalence in children living in rural area than their counterpart in the urban region. As observed by Kandala*et al* (2011), with reference tothe province of residence, the lowest prevalence ofstunting is observed in Kinshasa, the capital-city,whereas the highest is observed in provinces under warduring the survey (Equateur, Orientale, Nord-Kivu, Sud-Kivu and Maniema). The risk for a child living in theseprovinces to experience stunting is double of a child livingin Kinshasa.

Stunting in relation to the parent wealth index, the wealth index of the parents are grouped as Poorest, which is the reference (Wealth Index1), second (Wealth Index2), middle (Wealth Index3), fourth (Wealth Index4) and Richest (Wealth Index5). Wealth of the parents has negative relationship with the children stunting (Gaussian) in the sense that the richest parents have more stunting children of about 42% higher than the poorest parent children, the fourth have 8%, the middle and the second rich parents have 1% more stunting children than the poorest. While the richer the parents the less severely stunting the child, as the richest parent has 20%less severely stunting children, as well as the fourth with 11%less, the middle 7%less and the second rich parents with 4%less severely stunting children. Hence, severe stunted children is prevalence with poor parents as observed by Kandala*et al* (2011), that stunting is linearly associated withsocio-economic status of the household (higher amongchildren from the poorest household, followed by childrenfrom poor, middle or rich households but loweramong children from richest households: 49.8, 48.0,45.5, 43.9 versus 28.7 percent).

Mother education inversely influence the moderate stunting status of their children, as children from mothers with primary education have almost equal chance with children from mothers with no education. While mother with secondary school education and above have 14% more of moderate stunted children than none educated mothers, this was supported by the findings of Kandalla et al (2011), that there is no significant association between maternal education and the prevalence of stunting among children under the age of 5 years in the DRC. On the other hand, mother education has positive effect on severely stunted children, as mother with secondary education and above has 7%less of severely stunted children than children from non-educated mothers, with 11% more for children with primary education mothers. Therefore, the more educated the mothers the less severely stunted their children as reported by Emina*et al* (2011), that severe stunting is linearly associatedwith maternal education (higher among childrenfrom non-educated mother, followed by children frommothers with primary education but lower among childrenfrom mothers with secondary or higher education 49.8, 47.0 versus 35.2 percent).

Male children are 13%less stunted moderately than their female counterpart, while they (male) are more severely stunted by 11% than female. Previous research in this direction shows that the prevalence of stunting was higher among boyscompared to girls (46.1 versus 41.7 percent). And that stunting has aninverse linear association with the age of the child(higher in the age groups ranging from 4 years, followedby 3 years, 2 years, 1 years but lower in the younger age(0 year): 55.1, 49.4, 48.5, 46.5 versus 23.1 percent).

On the regional effect, the northern regions have less prevalence of moderate stunting children with North East and North West having 32% and 66% less respectively compared with the North Central, while the Southern regions have more prevalence of moderate stunted children with South East and South-South having 24% and 9% more respectively when compared with the North Central, South West is 0.3% less of moderate stunted children. On the other hand, the prevalence of severe stunted is pronounced in the Northern regions where the North East and North West were having 20% and 56% more of severe stunted children, while the Southern regions have less prevalence with South East, South-south and South West have 31%, 44% and 27% less of severe stunted children compared with the North Central.

The nonlinear effect of child's age in the Stunting Gaussian processis displayed in Figure 1a. The graph shows that the nutritional status of the child followed a downward slope from left to right, which implies that as the child grows the nutritional status is declining. That is, more children become stunted after two years of age. Hence the age of the child influences his nutritional status. Figures 1b, is map of Nigeria showing the posterior probabilities of significance estimates of the spatial effects. In Figure 1b, the colour white is associated with positively significant states, the colour black with negatively significant states, and the colour grey with non-significant states. The posterior means within 95% credible interval showing that Kano, Niger, Kwara, and Oyo states are states with more stunting children, while Gombe, Adamawa, Taraba, Plateau, and FCT Abuja having less stunting children, with the remaining states are not statistically significant in children with stunting. (Figure1).

While the Stunting Binomial regression shows that the age of the child (figure 2a) indicates an upward trend, although somehow irregular, which implies that the severely stunting children get improved as they grow. On the state performance (figure 2b) regardingsevere stunting, the 95% confidence interval shows that only Zamfara, Gombe, Taraba and Benue states have more pronounced severely stunted children, while Sokoto, Kebbi and Jigawa has less with the other state having non-significant effect of severe stunting children. (Figure 4).

In comparing the Gaussian and Binomial analysis, one important thing to note is that, Gaussian regression analyses assume a normally distributed data, the properties of a normal distribution holds. This implies that the Gaussian analysis result is for moderate or global nutritional deficiency status, while the Binomial analysis result is for severe cases of nutritional deficiency. Hence, the only condition for comparison is to see which of the determinants is moderate or severe with respect to which of the factors. For this reason, it means that the bases of their comparison would not be to infer that one method of the analysis is more suitable than the other, since the parameters are assessed with different perspectives.

5.0 Conclusion

The aim of site-specific province analysis is to accelerate policy interventions, optimise inputs (unobserved factors such as distal ones: food security and prices policies, environmental), improve child nutrition by taking into account the environmental impact and reduce the timescale to attain the Millennium Development Goals (MDGs). It is an approach that deals with multiple groups of factors input to improve child nutritional status in order to satisfy the actual needs of parts of the provinces rather than average needs of the whole country.This research work study childhood malnutrition which had been viewed with respect to stuntingand further grouped into moderate and severe condition with a view to have a thorough understanding of the specific nutritional status and to determine the effects of various factors such as place of residence, parents wealth index, mothers educational status, sex of the child and geographical location (the geo-political zones), while the child's age and states were considered as spatial effects.

References

Adebayo, SBA and Fahrmeir, L. (2002): Analyzing Child Mortality in Nigeria withGeoadditive Survival Models. Hieronymus, Munich.

Biggeri, A., Dreassi, E. and Marchi, M. (2004): A Multilevel Bayesian Model for contextual effect of material deprivation. Statistical Methods and Application, 13, 89-103.

Cox, D. R. (1972): Regression Models and Life Tables (with discussion). Journal of the Royal Statistical Society B, 34, 187-220.

Daniels, M. J. and Gatsonis, C. (1999): Hierarchical polytomous regression models with applications to health services research. *Statistics in Medicine***16,** 2311-2325.

Emina J.B, Kandala NB, Inung J and E.YYazoume (2011): Maternal Education and Child Nutritional Status in DRC. Journal of Public Health. 12, 576-592.

FAO (2004): Food and Agriculture Organization of the United Nations, Under Nourishment around the world. In the State of Food Insecurity in the World. Rome.

Federal Office of Statistics and UNICEF (1999): Multiple Indicator Cluster Survey 1999

Hastie, T.J. and Tibshirani, R.J (1990): **Generalized additive Model.** Chapman and Hall, New York.

Hennerfeind, A., Brezger, A. &Fahrmeir, L. (2006).Geoadditive survival models.Journal of the American Statistical Association, 101, 1065-1075.

Ibrahim, J.G., Chen, M.H. and Sinha, D. (2001); **Bayesian Survival Analysis**.Springer, New York.

Kandala NB, Fahrmeir L, Klasen S, Priebe J, (2009): Geo-additive models of Childhood Under nutrition in three Sub-Saharan African Countries. Population, Space and Place 2009, 15, 461-473.

Kandala NB, Madungu TP, Emina JB, Nzita KP, Cappuccio FP (2011). Malnutrition Among Children Under the Age of Five in the Democratic Republic of Congo (DRC): Does Geographic Location Matter? BMC Public Health, 1-15.

Kneib, T. and Fahrmeir, L. (2004): A mixed model approach for structured hazard regression. SFB 386 Discussion Paper 400, University of Munich.

Neal, R. M. (1996). **Bayesian learning for neural networks**. Lecture Notes in Statistics No. 118. New York: Springer-Verlag.

NDHS (2008): National Demographic Health Survey. Published by Federal Ministry of Health, Nigeria.

Oystein, K. and Ivy K. (2011): Children’s Stunting in Sub-Saharan Africa: Is there an Externality Effect of High Fertility? Demographic Research, 25, (18), 565-594.

Rasmussen, C.E. (1996): Evaluation of Gaussian Processes and other Methods for Non-Linear Regression. Ph.D. thesis. Graduate Department of Computer Science, University of Toronto.

SOWC, (2007): **State of the World Children**: A United Nations Children’s Fund (UNICEF) publication. Published by Oxford University Press, UK.

SOWC, (2011): **State of the World Children**: A United Nations Children’s Fund (UNICEF) publication. Published by Oxford University Press, UK.

UNICEF (2006): **Monitoring the Situation of Children and Women**. Multiple Indicator Cluster Survey Manual, New York.

UNICEF/FOS, (1999): **Multiple Indicators Clustering Survey**. New York. UNICEF, 1999

Williams, C.K.I. and C.E. Rasmussen (1996): **Gaussian processes for regression**. In: *Neural Information Processing Systems -8*. MIT press. Cambridge, MA. pp. 514–520.

World Health Organization. (2003): World Health report. Geneva. The Organization, 2003

Zellner, A. and Rossi, P. E. (1984) Bayesian analysis of dichotomous quantal response models.*Journal of Econometrics***25,** 365-393.

FIGURES

Fig.1a: Effects of Child Age (in Months)

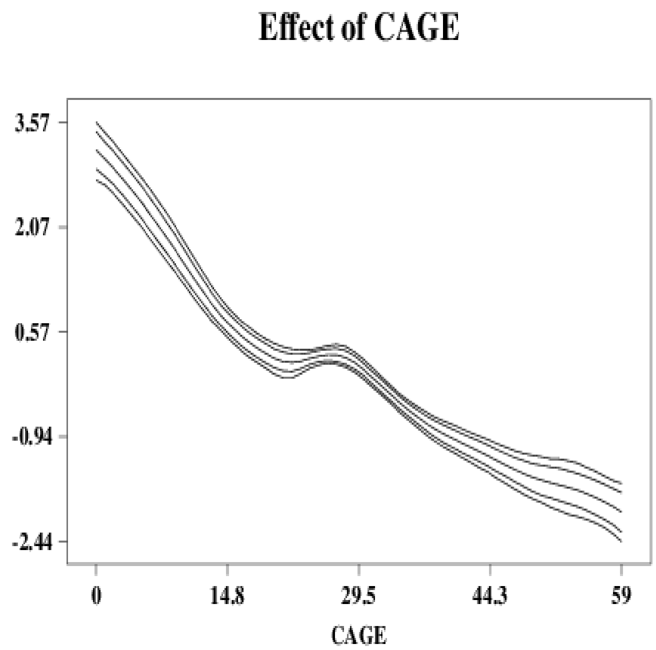

Fig.1b: State Effect on Stunting (Gaussian on Stunting (Gaussian Analysis) Analysis) at 95% CI

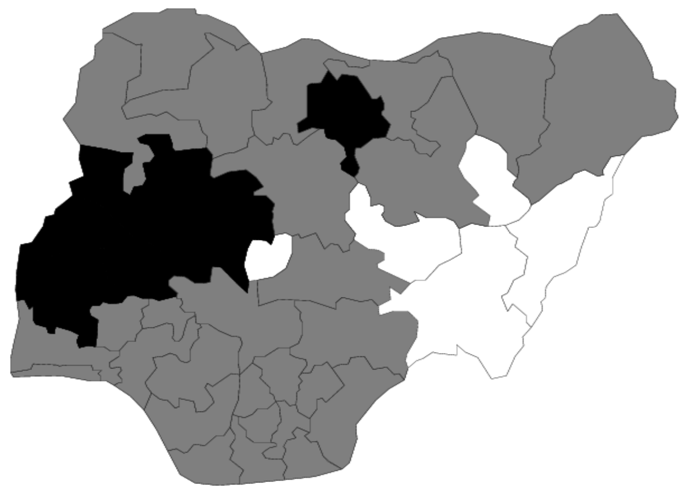

Fig.2a: Effects of Child Age (in months)

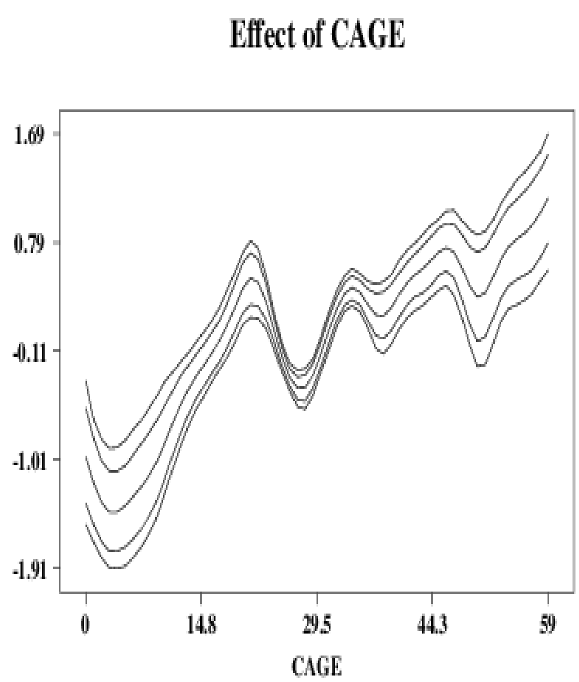

Fig.2b: State Effect on Stunting (Binomial on Stunting (Binomial Analysis) Analysis) at 95% CI