Modelling Immunisation Coverage in Nigeria: a Bayesian Structured Additive Regression Approach

Samson Babatunde Adebayo
Research and Evaluation Division, Society for Family Health, PMB 5116 Wuse, Abuja, Nigeria,
samson.adebayo@nafdac.org.ng; adebayosam@hotmail.com

Waheed Babatunde Yahya
Department of Medical Statistics and Epidemiology, Technical University of Munich, Germany,
wb_yahya@daad-alumni.de; wbyahya@unilorin.edu.ng

Abstract
About 10 million children under the age of five die every year worldwide. One quarter of these deaths are caused by diseases that are preventable with vaccines. According to the World Health Organisation, immunization currently saves between 2 and 3 million lives per year. It is one of the most successful and cost-effective public health interventions. In developing countries in particular, infant and childhood mortalities are related to childhood diseases. Therefore, low vaccination coverage increases the risks of a child to various diseases such as diarrhoea, measles, malaria, etc. In spite of the efforts from government and donor agencies, Nigeria still remains the county with least vaccination coverage in Africa. However, empirical evidence revealed substantial geographical variations on immunization coverage in Nigeria. In attempt to address the menace of low vaccination coverage in Nigeria, this paper aims at providing policymakers with tools to design effective interventions which can lead to frugal utilization of the scarce resources which is prominent in developing countries. This paper explores trend and geographical variations of vaccination coverage in Nigeria using 1999, 2003 and 2008 Demographic and Health Survey data. This approach permits a joint estimation of the usual linear effects of categorical covariates, nonlinear effects of continuous covariates and small-area district effects within a unified structured additive Bayesian framework.

Keywords: Deviance Information criterion, geo-additive regression, Markov chain Monte Carlo technique, preventable diseases, vaccination, Nigeria.

1 Introduction

The introduction and usage of different kinds of vaccines have contributed immensely to the eradication of some of the dreaded diseases in many developed countries. Immunization has remained the most cost-effective mechanism through which the outbreak of common diseases is prevented in many developing countries (Odusanya *et al*., 2000 & 2003). More than two million deaths among children are averted all over the world annually through vaccination against some early childhood diseases such as diphtheria, tetanus, pertussis, measles and hepatitis B (Duclos *et al*., 2009). This improvement in childcare notwithstanding, vaccine preventable diseases still contribute significantly to the global child mortality cases (Centre for Global Development, 2005). Particularly in 2002, the World health Organization (WHO) estimated under 5 mortality cases that are attributable to vaccine preventable diseases to be 1.4 million worldwide. This was about 14% of total global child mortality all over the world with considerable number of cases being from third-world countries.

In 1974, the Expanded Program on Immunization (EPI) was initiated by the World Health Assembly to assist in coordinating the efforts of public health programmes especially in developing countries to ensure full immunization of all children under one year of age against common diseases like poliomyelitis,

smallpox, diphtheria, tetanus, measles, tuberculosis, pertussis and so on. The EPI further aimed at ensuring that new vaccines and preventive health interventions are extended to children in parts of the world. One of the objectives of the EPI was that by 2010, global routine immunization coverage of all children under one year of age should reach 90% (Hadler *et al*., 2004). To ensure that the objectives of the EPI are achieved and sustained especially in poor countries of the world, the Global Alliance for Vaccines and Immunization (GAVI), a coalition of bodies such as United Nations Children Emergency Fund (UNICEF), WHO, and the World Bank was created in 1999 (Brugha *et al*., 2002). This body was enhanced by the establishment of Global Immunization Vision and Strategy (GIVS) (2006 – 2015) in 2005 at the 58th World Health Assembly. A common objective for establishing both GAVI and GIVS among others is to strengthen national immunization program and improve child and maternal health especially in the third world countries (Bilous *et al*., 2006).

We adopt a flexible Bayesian structured additive regression approach which permits joint estimation of trend, non linear effects of continuous covariates, geographical variations and fixed effects of categorical covariates. In the present study, we investigate the influence of bio-demographic variables such as maternal and partner (spouse) educational attainment, mother's age at the birth of child as well as some other socio-economic variables on vaccination coverage in Nigeria using flexible geoadditive models..

2 Data description

The datasets used for this study are the NDHS data for the three years 1999, 2003 and 2008. The Demographic and Health Surveys (DHS) are national representative surveys of men and women of reproductive age and their children in many developing countries of the world. These surveys are funded by *United States Agency for International Development* (USAID) for the purpose of collecting vital up-to-date information on health related matters such as mortality, morbidity, vaccinations and general health conditions of children and their mothers, as well as on many other socio-economic related variables that directly affect the growth and development of the children.

Information on vaccination coverage and on types of vaccines administered on the children through different immunization schedules are also included in all the NDHS data. The types of vaccines provided free by donor agencies to Nigerian children are *Bacille Calmette Guerin* (BCG) vaccines and *Oral Polio Vaccine* (OPV) to guide against tuberculosis and poliomyelitis respectively. Others are *Diphtheria, Pertusis, and Tetanus* (DPT) and measles vaccines. According to *Nigerian National Program on Immunization* (NNPI) schedule (which is adapted from the WHO immunization schedule), a child is considered to be fully vaccinated if he or she has received a BCG, three doses of DPT (i.e. DPT1, DPT2, DPT3), at least three doses of OPV (i.e. OPV0, OPV1, OPV2), and one dose of measles vaccines (NPC [Nigeria] & ICF Macro, 2009).

3 Structured Additive Regression Model and Analysis

3.1 Structured Additive Regression Model

Generalized linear models (e.g. Fahrmeir & Tutz, 2001) assume that, given covariates vector x and unknown parameters γ, the distribution of the response variable y belongs to an exponential family, with mean $\mu=E(y|\mathbf{x}, \gamma)$ linked to a linear predictor η by

$$\mu = h(\eta) \qquad \eta = \mathrm{x}'\gamma. \tag{1}$$

Here h is a known response function, and γ are unknown regression parameters. However, in most practical regression situations, we are often faced with the problem of rigid assumption of linear effect of continuous covariates in the datasets on the predictor. Sometimes, observations may be spatially or

temporally correlated. Furthermore, covariates may not be able to sufficiently describe any inherent heterogeneity among individuals or units. To overcome these difficulties, we replace the strictly linear predictor by a structured additive predictor.

Consider a set of regression observations (y_i, x_i, s_i, v_i), i=1, …, n, where y_i is either a binary or categorical response variable, a vector x=(x_1, …, x_p)' of continuous covariates (say respondents' age), s_i=(1,…, S) the state (district) where respondent i lived during the survey and a further vector v=(v_1, …, v_q)' of categorical covariates. Usually one intends to jointly model the dependence of y_i on continuous, spatial and categorical covariates within the context of generalized additive model (Hastie & Tibshirani, 1990).

3.2 Model Specification

Model A: A child is considered fully vaccinated if he or she has received a BCG vaccination against tuberculosis, three doses of DPT, vaccine to prevent diphtheria, pertussis, and tetanus; at least three doses of polio vaccine; and one dose of measles vaccine within the first year (NPC [Nigeria] & ICF Macro, 2009). A binary variable that describes level of vaccination coverage as

$$y_i = \begin{cases} 1: \text{if a child aged 12 months and beyond has received all the recommended vaccines} \\ 0: \text{otherwise} \end{cases}$$

was created.

Model B: A three-level ordinal outcome variable describing level of vaccination coverage was created as

$$y_i = \begin{cases} 1: \text{if a child aged 12 months and beyond has received all the recommended vaccines,} \\ 2: \text{if a child aged 12 months and beyond only received some recommended vaccinations,} \\ 3: \text{if a child aged 12 months and beyond did not receive any vaccination.} \end{cases}$$

The predictors in these models include non-parametric effect of a metrical covariate (mothers age at the birth of the child - measured in years, and child's age – measured in months), spatial components and linear part in an additive form. Similarly for model B, a cumulative probit model was assumed with the aim of modelling influence of determinants of level of vaccination coverage within a Bayesian perspective that jointly accounts for nonlinear, time, fixed and spatial effects in a similar manner as in model A.

3.3 Cumulative probit model

Let us consider the regression model based on multicategorical outcomes. Such models can be motivated from latent variables such that the response variable y can be observed in ordered categories 1, . . . , k. It is postulated that y is a categorized version of a latent variable

$$U = \eta + \varepsilon \tag{2}$$

obtained through the threshold mechanism

$$y = r, \qquad \theta_{r-1} < U \le \theta_r, \qquad r = 1,..., k$$

with thresholds $-\infty = \theta_0 < \theta_1 < \ldots < \theta_k = \infty$. We assume that the error variable ε has a distribution function F, hence it follows that y obeys a cumulative model

$$p(y \le r \mid \eta) = F(\theta_r - \eta). \tag{3}$$

To enhance identifiability, functions are centred about zero, thus the fixed effect parameters automatically include an intercept term. In the application to model *B*, level of vaccination coverage y is considered as a three-ordered categorical version of the latent continuous variable U. Here ε is assumed to have a standard normal distribution function, i.e.

$$p(y \le r \mid \eta = x, s, v) = \Phi(\theta_r - \eta)$$

yielding a cumulative probit model. Cumulative models based on category boundaries or threshold approaches (Edwards & Thurstone, 1952) are commonly used in ordinal regression.

5 Data Analysis and Discussions

5.1 Analysis

To explore impact of trend, demographic characteristics, continuous variables and spatial effect on level of vaccination coverage in Nigeria between 1999 and 2008, structured additive regression model was fitted. Model building was guided by the use of Deviance Information Criterion (DIC) proposed by Spiegelhalter *et al*. (2002). The following models were explored

M1: η = Trend alone
M2: η = Trend + Demographic characteristics
M3: η = Trend + Random effects (States and cluster) + Demographic characteristics
M4: η = Trend + Spatial effect (i.e. M1 + spatial)
M5: η = M3 + nonlinear effect of continuous variables
M6: η =Trend + Spatial + Random (States and Clusters) + Nonlinear of continuous + extended fixed effects (including partners' educational attainment)
M7: η =Trend + spatial + nonlinear of continuous + extended fixed effects (including partners' educational attainment)

It turned out that models with predictor M7 is the best in terms of the DIC. All analyses are carried out based on BayesX 2.0.1 - software for modelling structured additive regression modelling through a Bayesian perspective (Brezger, Kneib & Lang, 2009). This is available under http://www.stat.uni-muenchen.de/~bayesX

5.2 Results

Table 1 presents the results of the fixed effects while Figure 1 displays spatial effects for binomial probit and cumulative probit *models A* and *B* on map of Nigeria. The posterior means are shown in the left columns (a & c) while the corresponding posterior probabilities of significance of spatial effects are shown in the right columns (b and d).

6 Conclusions

This paper has provided readers with opportunity for flexibly modelling of nonlinear effects, spatial effects that incorporate neighbourhood influence, fixed effect and possibly random and interaction effects. In our analysis, we attempted random and interaction effects at an exploratory stage both found none significant. At another stage of the analysis, effect of continuous covariates i.e. child's age and mother's age at birth was assumed to be linearly related to *models A* and *B* and modelled parametrically. Even though these effects were significant in the parametric models, however, the model with smooth (nonlinear) functions of the covariates was found to be better in terms of the DIC. Evidently effects of child's age and mother's age at birth are non-linear, and an assumption of linear dependence *a priori* would have been too rigid and resulted in erroneous and spurious conclusions.

References

1. Bilous, J., Eggers, R., Gasse, F., Jarrett, S., Lydon, P., Magan, A., Okwo-Bele, J.M,, Salama, P., Vandelaer, J., Villeneuve, P. and Wolfson, L. J. (2006) "A new global immunisation vision and strategy," *Lancet*, 367,1464-1466.

2. Brezger, A., Kneib, T., & Lang, S. (2009) BayesX: Software for Bayesian Inference in Structured and Additive Regression Models, version 2.1. Available under: http://www.stat.uni-muenchen.de/~bayesx.
3. Brugha, R., Starling, M. and Walt, G. (2002) "GAVI, the first steps: lessons for the Global Fund," *Lancet*, 359,435-438.
4. Centre for Global Development (2005) *Making Markets for vaccines: from ideas to actions*, Centre for Global Development; Washington DC.
5. Duclos, P., Okwo-Bele, J-M, Gacic-Dobo, M., Cherian, T. (2009) "Global mmunization: status, progress, challenges and future," *BMC International Health and Human Rights*, 9 (Suppl. 1), S2 doi:10.1186/1472-698X-9-S1-S2.
6. Edwards D, Thurstone L. (1952) "An internal consistency check for scale values determined by the method of successive intervals," *Psychometrika*, 17,169–180.
7. Fahrmeir,L. and Tutz, G. (2001) "*Multivariate Statistical Modelling based on Generalized Linear Models*," Springer Verlag, New York.
8. Hadler, S., Cochi, S., Bilous, J., and Cutts, F. (2004) *Vaccination Programs in Developing Countries. Chapter 55:* Vaccines, Fourth Edition. Elsevier Inc.
9. Hastie, T., and Tibshirani, R. (1990) *Generalized Additive Models*, Chapman & Hall, London.
10. National Population Commission [Nigeria] and ICF Macro (2009) *Nigeria Demographic and Health Survey 2008*, National Population Commission and ICF Macro, Abuja.
11. Odusanya, O. O, Alufohai, J. E, Meurice, F. P., Clemens, R. and Ahonkhai, V. I. (2000) "Low immunization coverage in rural Nigeria," *Nigerian Quarterly Journal of Hospital Medicine*, 10,118-120.
12. Odusanya, O. O, Alufohai, J. E, Meurice, F. P., Clemens, R., Ahonkhai, V. I. (2003) "Short term evaluation of a rural immunization program in Nigeria," *Journal of the National Medical Association*, 95**,** 175-179.
13. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde A. (2002) "Bayesian Measures of Model Complexity and Fit (with Discussion)." *Journal of the Royal Statistical Society B*, 64, 583-640.

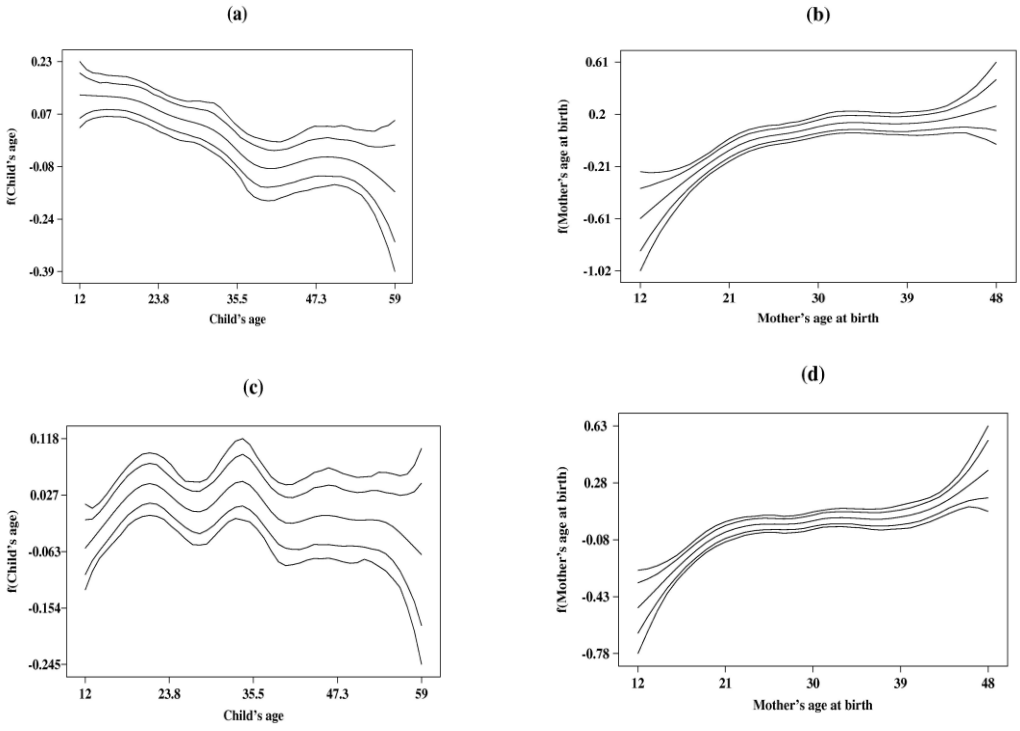

Figure 1: Nonlinear effects of (a) Child's age and (b) Mother's age at birth [Binomial probit model], (c) Chld's age and (d) Mother's age at birth with their corresponding 95% and 80% credible intervals.

Table 1: Posterior estimates for binomial and cumulative models with predictor M7.

Variables	Binomial model with predictor M7				Cumulative probit model with predictor M7			
	Mean	Std error	95% Credible Interval		Mean	Std error	95% Credible Interval	
			Lower	Upper			Lower	Upper
Constant	-2.160	0.254	-4.460	-3.441	-0.059	0.110	-0.279	0.149
Trend								
Year 1999 (ref)	ref				ref			
Year 2003	-0.141	0.143	-0.520	0.020	0.252	0.043	0.163	0.339
Year 2008	0.412	0.096	0.587	0.981	0.315	0.031	0.254	0.376
Place of residence								
Rural (ref)	ref				ref			
Urban	0.211	0.065	0.251	0.506	0.169	0.026	0.120	0.220
Sex								
Male (ref)	ref				ref			
Female	0.010	0.056	-0.079	0.138	0.028	0.021	-0.013	0.071
Respondents' educ								
None (ref)	ref				ref			
Primary	0.239	0.101	0.265	0.678	0.256	0.032	0.195	0.318
Secondary	0.419	0.107	0.591	0.993	0.520	0.040	0.442	0.604
Higher	0.572	0.138	0.758	1.281	0.694	0.062	0.577	0.812
Place of delivery								
At home/others	ref				ref			
Hospital	0.433	0.069	0.634	0.905	0.477	0.029	0.420	0.532
Threshold 1 (θ_1)	NA	NA	NA	NA	0.059	0.110	-0.149	0.279
Threshold 2 (θ_2)	NA	NA	NA	NA	2.234	0.112	2.018	2.463