

BAYESIAN HIERARCHICAL MODELING OF INFANT MORTALITY IN NIGERIA

OLUWAYEMISI OYERONKE ALABA AND CHIDINMA GODWIN

(Received 20 May 2019; Revision Accepted 14 June 2019)

ABSTRACT

Infant mortality and its risk factors in Nigeria was investigated using Bayesian hierarchical modeling. The hierarchical nature of the problem was examined to detect the within and between groups (states and regions) variations in infant deaths. The effect of individual level variables on the risk of a child dying before the age of one was determined using data collected from the fifth round Multiple Indicator Survey (MICS5, 2016-2017). Infants in Northern Nigeria had a higher risk of dying than others, especially in North West, while South West had the lowest risk of infant deaths. Ten percent of the variations in infant deaths was explained by differences between states while differences between regions explained only seven percent of the variations. Also, factors such as urban place of residence, mothers with secondary and tertiary education, first birth and birth interval above 2 years were associated with a decreased risk of infant deaths. Male infants, birth interval of less than 2 years, mothers with primary and no education, teenage mothers and mothers that gave birth at age 35 years and above were associated with a higher risk of infant mortality.

INTRODUCTION

Infant mortality is the demise of a child before age one which is an important index of the overall physical health of a country. Biologically, infants have weaker immune systems than adults which makes them susceptible to social or environmental complications (Caldwell, 1996). The determinant factors of infant mortality includes the level of education of the mother, environmental conditions, political and medical facilities, access to portable water, vaccination against contagious diseases, and other public health interventions. For over twenty years, infant and child survival has received global attention with a resultant effect of drastic decline in under-five and infant mortality rate (IMR) in some regions of the world (Claeson, 2000). Although, these achievements are challenged by differences that persist among countries.

The under developed and disadvantaged populations still grapple with the burden of infant mortality. In 2017, the global IMR was given at 29 deaths per 1000 live births, while it was 52 per 1000 live births in sub-Saharan Africa (UN, 2013).

The fifth round Multiple Indicators Cluster Survey (MICS5, 2016-2017) showed that IMR in Nigeria dropped to 70 per 1000 live births from 97 per 1000 live births in 2011. A report by UNICEF placed Nigeria in the 11th position in global ranking of infant mortality due to lack of assistance during delivery, poverty, conflict and weak institutions. Report also showed that although the situation has improved, the progress rate should be increased for Nigeria to attain the Sustainable

Development Goals.

Mosley and Chen (1984) developed a popular framework of the proximate causes of infant mortality which linked infant deaths to socio-economic determinants at individual, household, and community levels. Child's health and survival is hinged on the environment where she is born, especially in Africa (Entwisle et al. 2007, Arguillas 2008). The availability of health amenities in communities greatly influence the survival of young children (The Cebu Study Team 1991). Previous studies have increasingly used multilevel techniques to investigate the effect of contextual factors on infant mortality, as distinct from the more general individual factors. Socio-economic development, communities' educational and literacy levels have strong effect on infant's health outcomes and nutritional status (Parashar 2005, Boyle et al. 2006). Physical and social attributes of the place of residence also affects the health of young children, irrespective of the household context where the child is born (Arguillas 2008). Hence, this study is aimed at examining the variables that are associated with infant mortality, the direction and magnitude of the effect, and determination of the geographical variations (if it exists) in the occurrence of infant mortality in Nigeria. In this section we have discussed the prevalence of infant mortality globally and in Nigeria, and also some previous research that have been done in this area. In section 2, the estimation method employed is discussed. Results are presented in section 3, and discussion of results in section 4, while section 5 provides some concluding remarks.

Oluwayemisi Oyeronke Alaba, Department of Statistics, University of Ibadan, Nigeria.

Chidinma Godwin, Department of Statistics, University of Ibadan, Nigeria.

Data and Methods

Data

The study used data from Multiple Indicator Cluster Survey (MICS5, 2016-2017). The survey was conducted by the National Bureau of Statistics (NBS) in collaboration with the National Primary Health Care Development Agency (NPHCDA) and National Agency for the Control of Aids (NACA), as part of the global MICS programme. Two stage sampling technique was employed such that out of 37,440 households sampled, 35,747 households were visited, 34,289 were occupied and 33,901 were successfully interviewed, which gave a response rate of 98.9 percent. Hence, the data was collected from 33,901 households in 2,239 enumeration areas across Nigeria. A total of 34,376 eligible women, 28,085 of mothers/caregivers of children under 5 years, and 15,183 men were interviewed using structured questionnaires aided by Computer Assisted Personal Interview (CAPI) devices.

as:

$$Pr(y_i = 1) = \text{logit}^{-1} \left(\alpha + \sum_{i=1}^n \beta X_i \right) \quad \text{where } n = 5696 \quad (1)$$

$$\beta \sim N(\mu, \sigma^2) \quad (2)$$

$$\mu \sim N(0, 1000) \quad (3)$$

$$\sigma \sim U(0, 100) \quad (4)$$

One assumption of the single-level or standard logistic model is that the measured units are independent. If data are grouped and we have not taken account of group effects in our regression model the assumption will not hold.

Bayesian Hierarchical Logistic Model (Model 2)

In order to know if state and region of residence have significant effect on the risk of infant deaths in Nigeria and to determine the extent to which the effect of the individual variables vary between the clusters, we carried out a Bayesian hierarchical logistic regression analysis. This is the second model. Bayesian hierarchical modeling is a statistical model written in multiple levels that estimates the parameter of the posterior distribution using the Bayesian method. It is used when information is available on different levels of observational units and allows us to explore the

as:

$$Pr(y_i = 1) = \text{logit}^{-1} \left(\alpha_{jk[i]} + \sum_{i=1}^n \beta X_{ijk} \right) \quad \text{where } n = 5696 \quad (5)$$

$$\alpha_{jk} \sim N(\alpha_k, \sigma_{state}^2) \quad \text{for } j = 1, \dots, 37 \quad (6)$$

$$\alpha_k \sim N(0, \sigma_{region}^2) \quad \text{for } k = 1, \dots, 6 \quad (7)$$

where i , j and k are the indexes for the individual level, state and region respectively. The estimated group level variance represents the extent to which children in the same group (state or region) are exposed to the same conditions. This can be interpreted as the evidence of differential mortality levels between groups. The higher the estimated variance, the higher the level of inequality

METHODS

The outcome variable of interest is whether a birth in the last one year is still alive or not. We labelled the survey responses y_i as 1 for children that died and 0 for children that are still alive. The individual-level variables (X) considered are: type of place of residence, wealth quantile, mother's level of education, Mother's age at birth, gender, and previous birth interval.

Bayesian Logistic Model (Model 1)

The first model considered is the single-level Bayesian logistic regression to know the odds of an infant dying before his or her first birthday. It was carried out on the individual level variables as predictors, and does not take into account the possibility that a cluster factor might influence infant mortality. This is the base model. We assigned a normal prior to the β parameters and 2 their mean, while a uniform prior was assigned to the variance parameter. The model is expressed

geographical clustering present in the data. For this study state and region level predictors are not included.

Hence, a three-level logistic regression model with unobserved second and third level predictors was fitted on the survival status of children born one year before the census, where the first level is children below the age of 1, while the state and region in which the infant lives were the second and third level respectively. We assigned an inverse gamma prior to the state level variance parameter, while a half-Cauchy prior was assigned to the region level variance parameter because it has small number of groups (six). We used the same prior that were used for the standard Bayesian logistic regression for the individual level parameters.

The three level varying intercept logistic regression model can be expressed

estimated with Markov Chain Monte Carlo Methods (MCMC) using R2jags software version 0.5-7. The MCMC procedure is preferred because it gives less biased estimates of variance parameters than quasi-likelihood methods for binary response models (Browne, 2009). The Bayesian MCMC procedure is carried out by running the simulation for 1000 burn-in and 4000

INTRACLASS COEFFICIENT

Intraclass correlation coefficient (ICC) quantifies the proportion of observed data that is attributable to the effect of clustering (systematic differences between clusters). It can also be interpreted as the correlation

$$\rho_{0jk} = \frac{\tau_{0jk}}{u_{00k} + \tau_{0jk} + \sigma_y^2} \tag{8}$$

Where τ_{0jk} and u_{00k} are the residual variance at the second and third level respectively, while σ_y^2 is the level 1 error variance. In a multilevel logit model $\sigma_y^2 = \pi^2/3$ by assumption.

The ICC was calculated to quantify the degree of homogeneity of the outcome within clusters. It reflects the the unobserved factors that are shared among infants in the same group.

RESULTS

among observations within the same cluster, it ranges from 0 to 1, where ICC = 0 indicates that there is no between cluster variation, while ICC = 1 indicates that observations only vary between clusters. The usual formula for ICC for a three-level model is:

DESCRIPTIVE ANALYSIS OF VARIABLES

Table 1 shows the distribution of the explanatory variables and the survival status of infants in each geopolitical region in Nigeria. Among the 5696 infants included in the survey, 386 died before age one. 16.84% of these deaths occurred in North Central, 16.06% in North East, 40.93% in North West, 7% in South East, 8.29% in South South and 10.88% in South West. Table 2 shows the number and proportion of infants that died and those that are alive at different covariate levels.

Table 1: Descriptive analysis of dependent and explanatory variables

Table 1: Descriptive analysis of dependent and explanatory variables							
Variables	North Central freq (%)	North East freq (%)	North West freq (%)	South East freq (%)	South South freq (%)	South West freq (%)	Tota freq (%)
Survival status							
Alive	920(17.33)	896(16.87)	1874(35.29)	470(8.85)	598(11.26)	552(10.40)	5310
Dead	65(16.84)	62(16.06)	158(40.93)	27(7.00)	32(8.29)	42(10.88)	386
Residence							
Rural	728(17.49)	797(19.15)	1634(39.26)	386(9.27)	449 (10.79)	168(4.04)	4162
Urban	257(16.75)	161(10.50)	398(25.94)	111(7.24)	181(11.80)	426(27.77)	1534
Wealth index							
Poor	409(16.35)	601(24.03)	1307(52.26)	64(2.56)	63(2.52)	57(2.28)	2501
Middle class	243(21.58)	168(14.92)	347(30.82)	115(10.21)	162(14.39)	91(8.08)	1126
Rich	333(16.09)	189(9.13)	378(18.27)	318(15.37)	405(19.58)	446(21.56)	2069
Mothers edu							
None	370(14.12)	625(23.86)	1511(57.67)	23(0.88)	21(0.80)	70(2.67)	2620
Primary	231(26.19)	113(12.81)	212(24.04)	78(8.84)	147(16.67)	101(11.45)	882
Sec and higher	384(17.50)	220(10.03)	309(14.08)	396(18.05)	462(21.06)	423(19.28)	2194
Mothers age							
Less than 20	111(14.21)	163(20.87)	388(49.68)	30(3.84)	52(6.66)	37(4.74)	781
20-34	721(18.29)	646(16.39)	1291(32.75)	366(9.28)	481(12.20)	437(11.09)	3942
Above 34	153(15.73)	149(15.31)	353(36.28)	101(10.38)	97(9.97)	120(12.33)	973
Gender							
Female	494(17.24)	487(16.99)	1012(35.31)	246(8.58)	332(11.59)	295(10.29)	2866
Male	491(17.35)	471(16.64)	1020(36.04)	251(8.87)	298(10.53)	299(10.57)	2830
Birth interval							
First birth	190(19.13)	140(14.10)	296(29.81)	106(10.68)	130(13.09)	131(13.19)	993
Less than 2yrs	160(15.83)	179(17.70)	371(36.70)	117(11.57)	100(9.89)	84(8.31)	1011
2 years	316(18.03)	306(17.46)	631(35.99)	152(8.67)	200(11.41)	148(8.44)	1753
Above 2yrs	319(16.45)	333(17.17)	734(37.86)	122(6.29)	200(10.32)	231(11.91)	1939

* freq - frequency
* Mothers edu - Mothers educational attainment

Table 2: Distribution of infant mortality based on selected covariates

Variable	Alive Number (%)	Dead Number (%)	Total
Residence			
Rural	3850 (99.3)	117 (7.5)	4109
Urban	1460 (95.18)	74 (4.82)	1534
Wealth index			
Poor	2114 (92.52)	167 (7.48)	2501
Middle class	1029 (91.39)	97 (8.61)	1126
Rich	1907 (95.09)	102 (4.97)	2009
Mothers edu			
None	2407 (91.87)	213 (8.13)	2620
Primary	620 (93.97)	42 (7.03)	662
Sec or higher	2082 (94.94)	111 (5.06)	2193
Mothers age			
Less than 20	705 (94.27)	46 (9.73)	751
20-34	3700 (94.06)	230 (5.94)	3930
Above 34	890 (92.00)	77 (7.91)	967
Gender			
Female	2694 (94.00)	172 (6.00)	2866
Male	2616 (92.43)	214 (7.56)	2830
Birth interval			
First birth	926 (93.25)	67 (6.75)	993
Less than 2yrs	898 (88.82)	113 (11.18)	1011
2 years	1641 (93.61)	112 (6.39)	1753
Above 2yrs	1845 (95.15)	94 (4.85)	1939

= Mothers edu - Mothers educational attainment

BAYESIAN MODEL

Table 3 shows the model outputs. Both models gives the same direction of the effect of the variables considered, but the magnitude of the effect vary for some variables while it remains unchanged for others. For the multilevel model, infants residing in an urban area had 14% significant lower odds [OR: 0.86, CI: 0.74, 0.99] of dying compared to those in rural areas, also children from rich households had 7% lower risk of dying compared to those from poor households although not significant. Birth interval of more than 2 years significantly reduced the odds of an infant dying by 29% compared to birth interval of 2 years, while birth interval of less than 2 years was significantly associated with 59% higher odds [OR: 1.59, CI: 1.32, 1.92] of infant deaths. Male infants had 14% significant higher odds of dying than female infants. Mothers with no education was associated with 19% significant higher odds of infant deaths compared to mothers with secondary or higher education for model 1 but we found this effect to be insignificant for model 2, while the significant reduction in the risk of dying for

infants residing in urban areas compared to those in rural areas that was observed in model 2 was insignificant in model 1. The effect of mothers age, gender and birth interval is the same for both models. We can also see from table 3 that 7% of the variations in infant deaths is significantly explained by between regions differences, while 10% of the variations is significantly explained by between states differences.

To see the geographical variations in infant mortality in Nigeria, we look at the distribution of the predicted log odds of infant mortality of the states and regions. It can be observed from Figure 1 that the North West had the highest log odds of infant deaths in Nigeria, while the South West had the lowest log odds of infant deaths. Generally, the Northern parts of Nigeria seems to be associated with a higher log odds of dying than other regions. Figure 2 shows that among the states in Nigeria, Lagos and Edo state had the lowest log odds of infant deaths while Kano state had the highest log odds of infant deaths.

Table 3: Odds ratio and 95 % credible intervals for the effect of individual level factors on infant mortality in Nigeria

Variable	Model 1		Model 2	
	OR	95% CI	OR	95% CI
Individual level variables				
<i>Type of residence:</i>				
rural (ref)				
urban	0.87	[0.76, 1.01]	0.86	[0.74, 0.99]
<i>Wealth index:</i>				
poor (ref)				
middle class	1.21	[1.04, 1.42]	1.19	[1.03, 1.40]
rich	0.92	[0.76, 1.10]	0.93	[0.77, 1.12]
<i>Mother's education:</i>				
secondary or higher(ref)				
none	1.19	[1.01, 1.38]	1.16	[0.98, 1.38]
primary	1.05	[0.88, 1.25]	1.06	[0.88, 1.26]
<i>Mother's age:</i>				
20-34(ref)				
less than 20yrs	1.14	[0.95, 1.36]	1.14	[0.95, 1.37]
above 34yrs	1.12	[0.94, 1.33]	1.12	[0.93, 1.32]
<i>Gender:</i>				
female(ref)				
male	1.14	[1.03, 1.26]	1.14	[1.04, 1.27]
<i>Birth interval:</i>				
2years(ref)				
first birth	0.96	[0.79, 1.17]	0.96	[0.78, 1.18]
less than 2yrs	1.59	[1.33, 1.91]	1.59	[1.32, 1.92]
above 2yrs	0.71	[0.60, 0.84]	0.71	[0.59, 0.85]
σ_{state}			0.38	[0.26, 0.54]
σ_{region}			0.28	[0.01, 1.32]
DIC	2758.5		2758.4	
ICC state			0.10	
ICC region			0.07	

+ OR - Odds Ratio

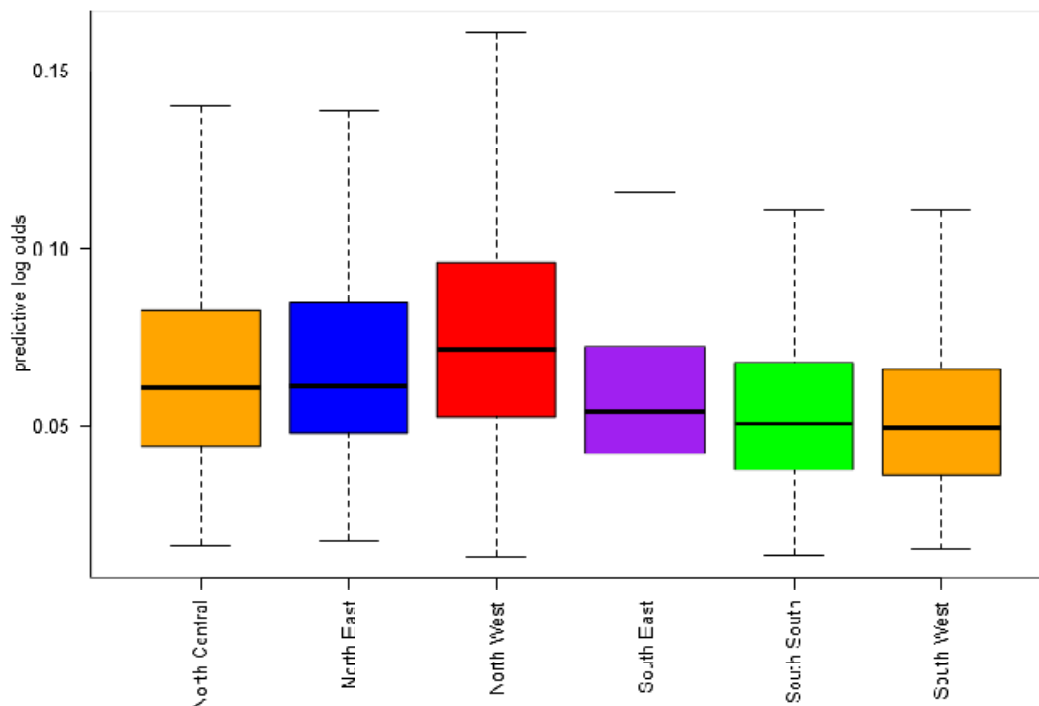


Figure 1: Boxplot showing predicted log odds of infant deaths across the geopolitical regions in Nigeria.

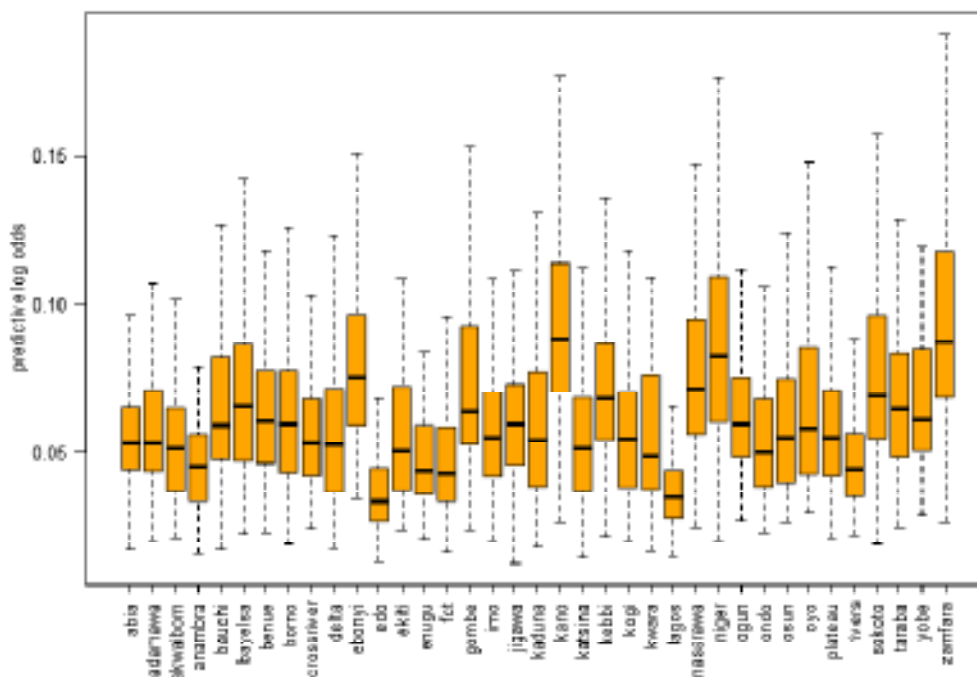


Figure 2: Boxplot showing predicted log odds of infant deaths across the states in Nigeria. 10

DISCUSSION OF RESULTS

The effect of place of residence, maternal age at birth and level of education, child's gender, household wealth quantile, and previous birth interval on infant mortality in Nigeria was examined. A Bayesian hierarchical model was employed to examine the influence of the states and regions of residence on an infant's risk of dying. The results indicated residual state and regional-level effect on the risk of infant deaths, even after controlling for individual-level explanatory variables. There was observed higher clustering at the state-level in comparison to the region-level. These results suggest that states and regions of residence also had a significant effect, though not to a large extent, on the risk of infant deaths. We found that infants in the North are at more risk of dying than infants in other regions. Type of residence, child's gender, and birth interval were significant determinant of infant mortality. The estimates obtained from the analysis showed that residing in an urban area reduced the risk of infant deaths in comparison to rural areas, this could be due to a better access to health care, portable water, and other improved environmental facilities enjoyed by people in urban areas. Birth interval above two years was found to be linked with a lower risk of infant deaths, while an interval below 2 years was related with a higher infant mortality risk compared to a birth interval of two years. This agrees with previous reports that close child spacing increases infant mortality risk (Bolstad and Manda 2001) possibly due to maternal depletion syndrome and contest between siblings for available resources (Rutsein 2005). It was also observed that male infants have higher mortality risk than female infants. This finding seems to be in agreement with reports that male infants are biological disadvantaged in Africa (Waldron, 1998). Before accounting for state-level and region-level differences, mother's with no education

mortality while the effect of residing in an urban area was insignificant, but after accounting for these differences, the effect of no maternal education was found to be insignificant while that of place of residence was significant, suggesting that not accounting for the geographical clustering resulted in incorrect estimated effects. A major surprise is the finding that middle class households had higher risk of infant deaths compared to poor households. This observation seems controversial and requires further investigation for a better understanding of this reversal from previous reports.

CONCLUSION

The study analysed infant mortality in Nigeria by considering the hierarchical nature of the data and investigated the geographical variation of infant mortality. The result suggests that child spacing through family planning and providing improved facilities in the country can reduce infant mortality. The results from the hierarchical modelling shows that there is need for state and geopolitical region level implementation of policies to drastically reduce infant mortality, especially in states with higher prevalence.

REFERENCES

- Arguillas, M. 2008. A Multilevel Analysis of the Community and Household Determinants of Child Mortality in Metro Cebu, the Philippines. Ph.D. Thesis, Cornell University: NewYork, United States.
- Bolstad, W.M and S.O. Manda. 2001. Investigating Child Mortality in Malawi Using Family and Community Random Effects: A Bayesian Analysis. *Journal of the American Statistical Association* 96: 12-19(8).

- Caldwell P., 1996. Child survival: Physical vulnerability and resilience in adversity in the European past and the contemporary Third World. *Social Science and Medicine*, 43(5): 609. [https://10.1016/0277-9536\(96\)00109-8](https://10.1016/0277-9536(96)00109-8).
- Boyle, M.H., Y. Racine, K. Georgiades, D. Snelling, S. Hong, W. Omariba, P. Hurley, and P. Rao-Melacini. 2006. The Influence of Economic Development Level, Household Wealth and Maternal Education on Child Health in The Developing World. *Social Science and Medicine* 63: 2242-2254.
- Claeson M, Bos E, Mawji T, Pathmanathan I. Reducing child mortality in India in the new millennium. *Bull World Health Organ*. 2000; 78:1192–9. Pub Med Central Pub Med.
- Entwisle, B., K. Faust, R.R. Rindfuss, and T. Kaneda. 2007. Networks and Contexts: Variation in the Structure of Social Ties. *American Journal of Sociology* 112: 1495-1533.
- Forste, R. 1994. The Effects of Breastfeeding and Birth Spacing on Infant and Child Mortality in Bolivia. *Population Studies* 48: 497-511.
- Griffiths, P., N. Madise, A. Whitworth, and Z. Matthews. 2004. A Tale of Two Continents: A Multilevel Comparison of the Determinants of Child Nutritional Status from Selected African and Indian Regions. *Health and Place*. 10: 183-199.
- Hobcraft, J.N., J.W. McDonald, and S.O. Rutstein. 1985. Demographic Determinants of Infant and Early Child Mortality: A Comparative Analysis. *Population Studies* 39: 363-385.
- Kravdal, O. 2004. Child Mortality in India: the Community-Level Effect of Education. *Population Studies* 58:177-92.
- Ladusingh, L. and H.C. Singh. 2006. Place, Community Education, Gender and Child Mortality in North-East India. *Population, Space and Place* 12: 65-76.
- Lawn J.E, Cousens S, Zupan J. Lancet Neonatal Survival Steering Team: 4 million neonatal deaths: When? Where? Why? *Lancet*. 2005; 365:891–900.
- Manda, S.O.M. 1988. Unobserved Family and Community Effects on Infant Mortality in Malawi. *Genus* 47: 1841-1854.
- Mosley, W.H. and L.C. Chen. 1984. An Analytical Framework for the Study of Child Survival in Developing Countries. *Population and Development Review* 10: 25-45.
- Parashar, S. 2005. Moving beyond the Mother-Child Dyad: Women's Education, Child Immunization, and the Importance of Context in Rural India. *Social Science and Medicine* 61: 989-1000.
- Rutstein, S.O. 2005. Effects of Preceding Birth Intervals on Neonatal, Infant and Under-Five Years Mortality and Nutritional Status in Developing Countries: Evidence from the Demographic and Health Surveys. *Int J Gynaecol Obstet* 89 Suppl 1:S7-24.
- The Cebu Study Team. 1991. Underlying and Proximate Determinants of Child Health: The Cebu Longitudinal Health and Nutritional Study. *American Journal of Epidemiology* 133: 185-201.
- UN Inter-agency Group for Child Mortality Estimation. Levels and trends in child mortality. Report 2013.
- Waldron, I. 1998. Sex Differences in Infant and Early Childhood Mortality: Major Causes on Death and Possible Biological Causes. Pp. 64-83 in *Too Young to Die: Genes or Gender?* New York (US): United Nations.
- WHO (2018): Global Health Observatory Data. Child mortality and causes of death

APPENDIX

Model Diagnostics

To check for convergence of the posterior distribution, we use trace plots. Trace plot shows the parameter estimates obtained from successive iterations of the

chains. If the chains converge, the plot should show rapid up-and-down variation with no trends. The trace plots and marginal density plots in Fig 3 shows that the chains converges for all the parameters represented by $b[1], \dots, b[11]$.

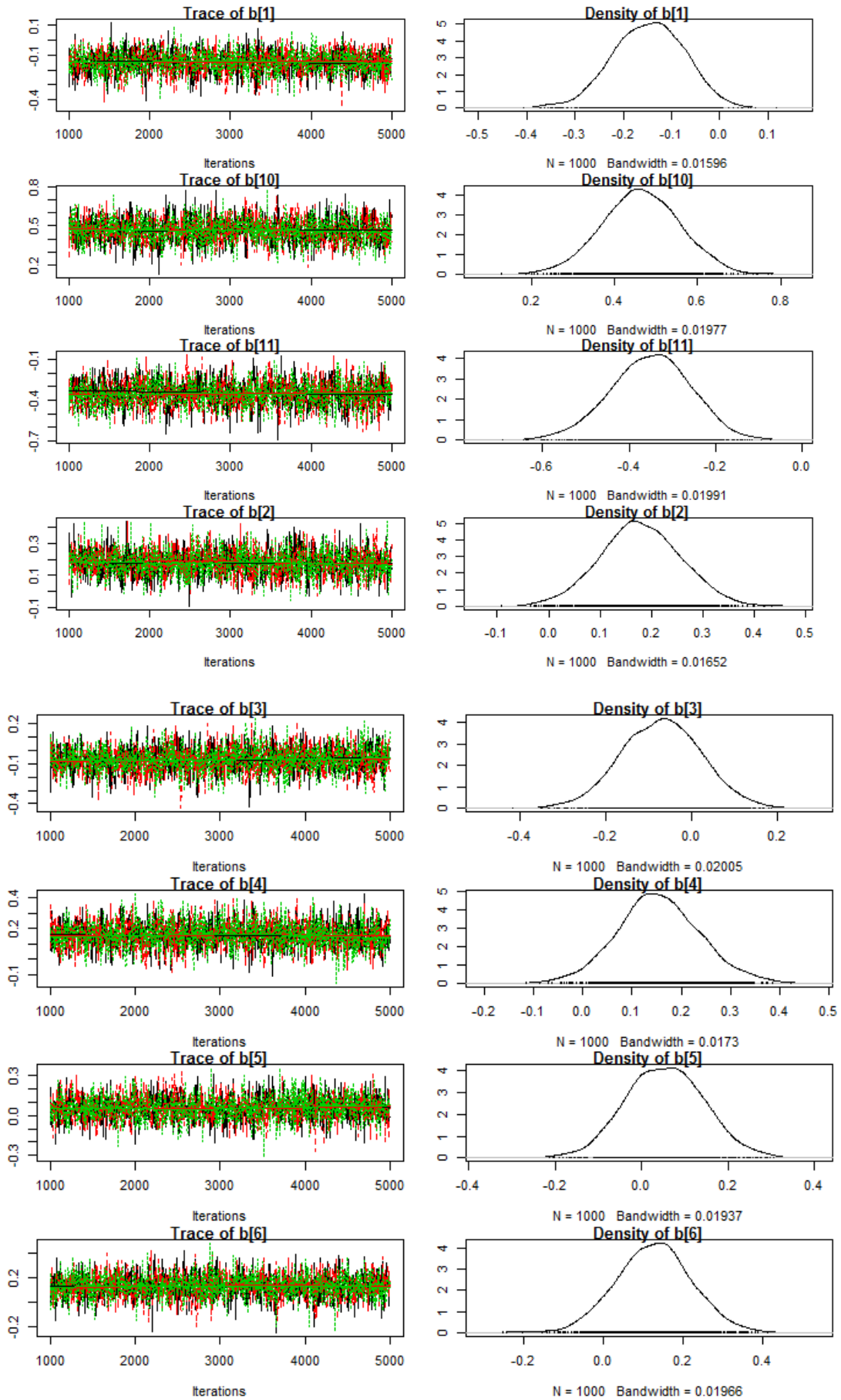


Figure 3: Traceplot showing convergence.

Posterior predictive checking was used to assess the fit of the model. Datasets were generated from the predictive distribution of the model, and were compared to the observed datasets. This goodness-of-fit tests were conducted using DHARMA R-package. Fig 4 shows the QQ plot on the left and residual plot on the right. QQ-plot detects overall deviations from expected

distribution and also test for uniformity and outliers. The p-values obtained from the Kolmogorov-smirnov (KS) test and the QQ-plot showed that the residuals follow the expected uniform distribution, the residual plot also showed that there was no pattern in the residuals. The result of the outlier test implies that there are no significant outliers.

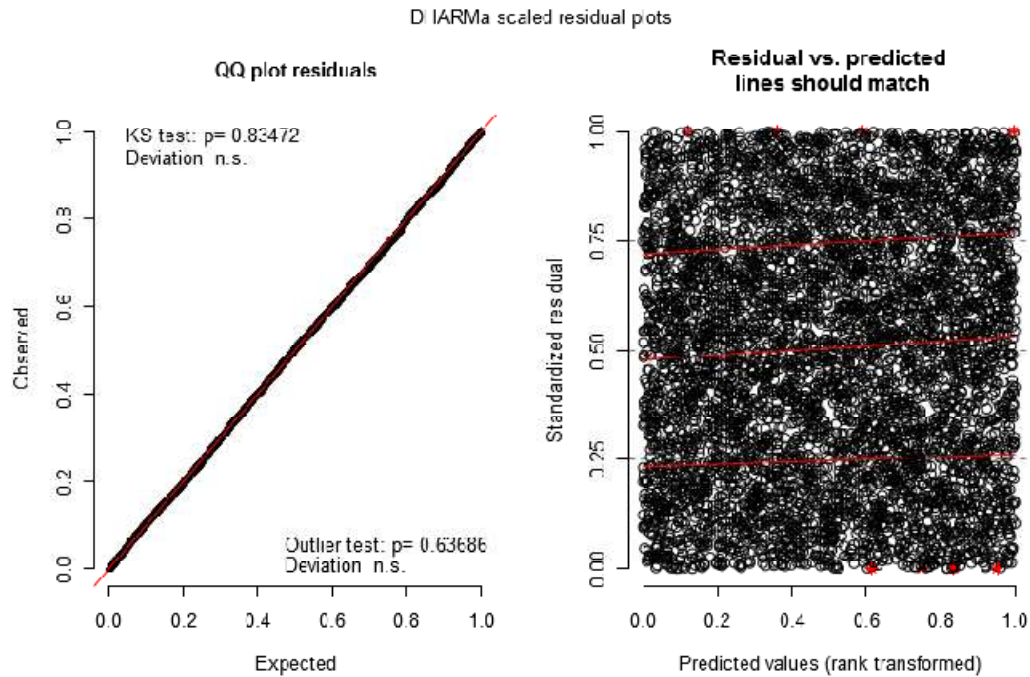


Figure 4: Residual plot.

Note: When interpreting DHARMA residuals, the residuals are expected to follow a uniform distribution instead of the normal distribution, and are standardized to values between 0 and 1. Fig 5 shows that there is no over/under dispersion or zero-inflation in the model. Over-dispersion occurs when the observed variance is

higher than the variance of a theoretical model, while under dispersion means that that the observed data has less variation than the predicted data. Zero inflation occurs when the observed proportion of zeros is greater than expected from the theoretical logistic model.

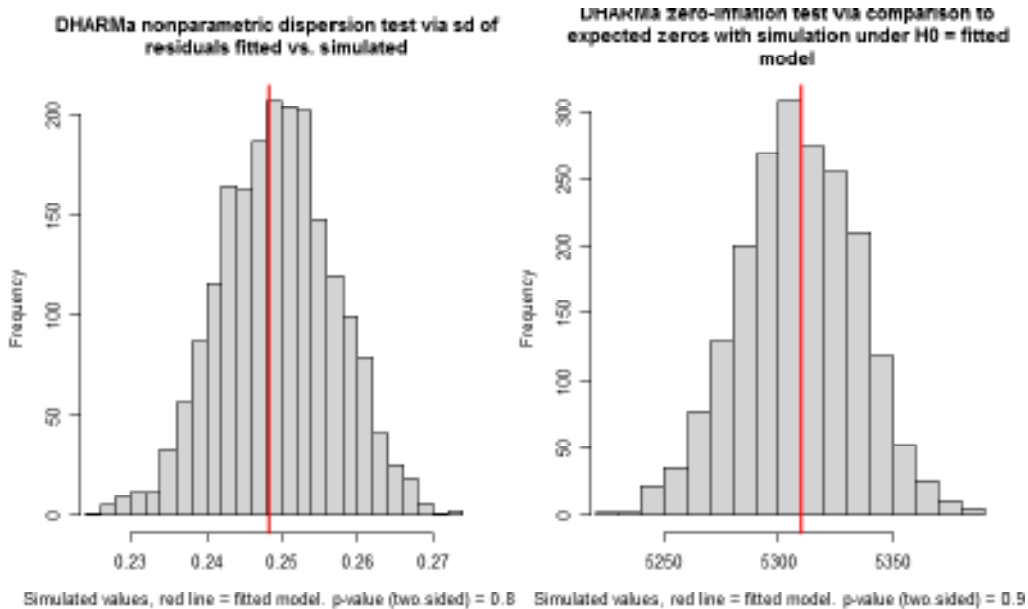


Figure 5: Tests for dispersion and zero-inflation.