FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 3 No. 4, December, 2019, pp 328 –344

A BAYESIAN SIMULATION MODELING APPROACH TO PREDICTING MATERNAL AGE-SPECIFIC INFANT SURVIVAL OUTCOMES WITH INCOMPLETE DATA

***Agada, P. O., Udoumoh, E. F. and Gboga, H. P**

Department of Mathematics/Statistics/Computer Science, University of Agriculture Makurdi, Benue State, Nigeria

*Corresponding Author Email: gadexx@yahoo.com, (+234 - 8038637735)

ABSTRACT

Predicting infant survival rates using the Classical Binary Logistic Regression Model with maternal and child characteristics as covariates can be a challenge when the modeler requires a Maternal Age-Specific model but that is not forthcoming. Reason being that mother age is not a significant covariate in the model. One way out of this is to group mother age at child birth into class intervals of age groups and see whether infant survival outcomes vary significantly with the age groups. If this is true, then Classical Binary Logistic Regression Models one for each age group, can be fitted for predicting infant survival outcomes. A fresh challenge would be that of incomplete data since the data set would have been merged by the age groupings. This new challenge can be overcome by the Bayesian Simulation Modeling Approach. Hence our task in this study is to develop a Bayesian Simulation Modeling Procedure implemented on the Simulation package; Windows Bayesian Inference Using Gibbs Sampling (WINBUG) with the aim of modeling the relationship between Infant Survival Outcomes and Maternal and Child characteristics, for each maternal age group. Besides the successful model fit in the face of incomplete data, the overall result of the study revealed that, the three maternal age groups; 15 – 25 years, 26 – 35 years and 36 years and above have positive impact on infant survival rate, while only the weight of infants delivered by mothers who are 36 years and above pose as risk factor to infant survival rate.

Keywords: Modeling, Logistic, Regression, WINBUG

INTRODUCTION

Research has shown that mother age has become a significant contributor to infant mortality rate in recent times. Selemani, M., Mwanyangala, M.A., Mrema, S. (2014) stated that the results from a logistic regression model indicated increase in risk of neonatal mortality among neonates born to young mothers aged 13–19 years compared with those whose mother‘s aged 20–34 years. Predicting infant survival rates using the Classical Binary Logistic Regression Model with maternal and child characteristics as covariates can be a challenge when the modeler requires a Maternal Age-Specific model but that is not forthcoming. Reason being that mother age is not a significant covariate in the model. One way out of this is to group mother age at child birth into class intervals of age groups and see whether infant survival outcomes vary significantly with the age groups. If this is the case, then a Classical Binary Logistic Regression Model fitted for each age group, can be used for predicting infant survival outcomes. It is important to mention that this will not be without the challenge of incomplete data, as the original data set would have been merged by the age groupings. Hence our task in this study is to develop a Bayesian Simulation Modeling Procedure implemented on the Simulation package; Windows Bayesian Inference Using Gibbs Sampling (WINBUG) with the aim of modeling the relationship between Infant Survival Outcomes and Maternal and Child characteristics, for each maternal age group.

Some authors support the above argument about the limitation of the use of the Classical Binary Logistic Model with incomplete data. One of such authors is Peduzzi et al., (1996) who emphasized the use of adequate data when the Classical Binary Logistic Model is been employed. This they did by suggesting a formula for computing the minimum sample size required, given the number of covariates and smallest value of the success rate. Taeryon et al., (2008) tried to remove this dilemma by suggesting the use of the Bayesian Simulation Approach to Logistic Regression Modeling where only aggregate Bernoulli trials are available and not the individual trials; what they termed incomplete data. Tripathi et al., (2019) developed a model for assessing child mortality under different parity using a Bayesian swatch. Other works in this direction include those of Gemperil, (2004) and Koissi and Hogens, (2005) to mention a few.

We implement this approach on a data set of Infant Survival Outcomes (alive or dead) and some maternal and child characteristics (Mother age at child birth and HIV Status, Infant sex and weight). This data set was sourced from the hospital records of the Madonna Hospital, Makurdi Benue State Nigeria. The data set was found suitable for the scenario earlier described.

The rest of the paper is organized as follows; Methodology, Results, Discussion, Conclusion and Recommendation.

MATERIALS AND METHOD

Model description

Binary logistic regression is a statistical modeling procedure for building regression models used for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two outcomes). The main objective of binary regression modeling is to find a model that best describe the existing relationship between the dichotomous characteristic of interest and the set of independent or predictor variables. The model achieves this by generating the coefficients of a formula to predict a logit transformation of the probability of the presence of the characteristic of interest:

$$Logit(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K \quad (1)$$

It follows from equation (1) that the Infant Survival Rate (%) can be modeled as

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_K X_K)}} * 100\% \quad (2)$$

Where p is the probability of the presence of the characteristic of interest, the β_i's and the $X_i's$, $i = 1,2,\ldots,k$ are the regression coefficient and the independent variables respectively. The regression coefficients or parameters can be estimated by the Method of Maximum Likelihood (MLE).

In this work, p is the probability of infant survival or infant survival rate, the independent variables include, mother age at child birth, mother HIV Status, child weight and child sex.

The logit transformation is defined as the logged odds where;

$$Odds = \frac{p}{1-p} = \frac{probability\ of\ the\ presence\ of\ the\ characteristic}{probability\ of\ the\ absence\ of\ the\ characteristic} \quad (3)$$

And

$$Logit(p) = Ln\left(\frac{P}{1-P}\right) \quad (4)$$

Unlike the ordinary regression model that chooses model parameters that minimize the sum of squares of errors, logistic models choose parameters that maximize likelihood of observing the sample values. A total of 974 cases were used in this work which according to Peduzzi et al., (1996) is considered adequate. The Forward Stepwise (Likelihood Ratio) regression method is employed in the modeling process. The fitted Binary Logistic Model which was well validated (with a good Hosmer – Lemeshow goodness of fit test result(p > 0.05 and overall percentage model correct prediction of 95% survival outcomes, see table 2 and table 3), retained weight of infant as the only covariate. Hence the covariate mother age at child birth, which the researcher looks out for is not included in the model. This led to chi - square test of homogeneity of the distribution of infant survival outcomes across three classes of age group of mothers at child birth. The three age groups identified were; 15 – 25, 26 – 35 and greater than or equal to 36 years.

Chi square test of homogeneity

The chi – square test of homogeneity is used to determine whether frequency counts are identically distributed across different populations or across different sub-groups of the same population. In this work, we test the homogeneity of the distribution of infant survival outcomes across the classes of age group of mothers. The null hypothesis is that the distribution of infant survival outcomes is the same across these three age groups. If this is true, then equal frequencies should be expected across the age groups. The chi – square statistic helps to determine whether the null hypothesis should be accepted or not. The chi- square calculated value is computed using the formula;

$$\chi_{cal}^{2} = \sum_{i}^{r} \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

Where O_i = observed or actual number of survivals, E_i = expected number of survival for the i^{th} age group and r = the number of age groups.

We have enough evidence to reject the null hypothesis at α level of significance if the critical value of the $\chi_{crit}^{2}(\alpha, r-1) < \chi_{cal}^{2}$. The result on table 5 shows that the null hypothesis is to be rejected. The implication is that, the distribution of Infant Survival Outcomes is Age Group Specific. Hence, Binary Logistic Models, one for each age group should be fitted but not without the challenge of sample size inadequacy.

Model sample size adequacy

The complexity in determining optimal sample size for logistic regression modeling made Peduzzi et al., (1996) suggested that, if p is the smallest of the proportions of positive or negative cases in a population and k the number of covariates, then the minimum number of cases to include is:

$$N = \frac{10 * K}{P} \quad (6)$$

If the resulting $N < 100$, Long (1997) suggested that it should be rounded up to 100.

If we consider p = 0.1 as the least proportion of infants who survived (the highest been 0.9 approximately see table 4) and four (4) independent variables, a sample size of 400 would be required. This is against the 307, 267 and 174 available cases for the three respective age groups. Hence, these sample sizes can be considered adequate for fitting Age Group Specific Classical Binary Logistic Models.

Model goodness of fit and adequacy checks

In order to test for the goodness of fit of the Classical Binary Logistic model, Hosmer – Lemeshow test was employed. The test divides the test data into approximately 10 groups. The chi-square statistic for this test is computed by;

$$\chi^2_{HL} = \sum_{g=1}^{G} \frac{(O_g - E_g)}{E_g(1 - E_g/n_g)} \quad (7)$$

with O_g, E_g and n_g defined as the observed events, expected events and number of observations for the g[th] decile group and G the number of groups. The number of degree of freedom is G-2. A large value of chi-square with small p-value < 0.05 indicates poor fit while a small chi-square vale with p –value closer to 1 indicate a good logistic regression model fit.

In order to evaluate the prediction accuracy of the Binary logistic model, the classification table is employed. On this table, the observed values of the dependent variable and the predicted values at a user defined cut - off value are cross classified.

The Walds statistic tests the significance of model parameters. This helps to determine whether or not an independent variable stays in the model as it tests if the associated model parameter differs significantly from zero. The Walds statistic is computed as the regression coefficient divided by its standard error squared:

$$\left(\frac{\beta}{SE}\right)^2 \quad (8)$$

Where $\beta = \text{the associated regression parameter}$, $SE = \text{It'standared error}$.

If the p-value is less than the usual $\alpha = 0.05$, then we have evidence that to conclude that the independent variable differ significantly from zero. Hence it stays in the model. This test was used to determine if the significant independent variable in the model, most especially whether infant weight was retained is significant.

Odds ratio

Re-writing equation (1) by taking the exponential of both sides of the equation, we have;

$$Odds = \frac{p}{1-p} = e^{\beta_0} e^{\beta_1 X_1} e^{\beta_2 x_2} \ldots e^{\beta_k X_k} \quad (9)$$

It is obvious from equation 5 above that when an independent variable X_i changes by 1 unit (all other variables kept constant), the odds changes by the factor e^{β_i}. This factor is termed the odds ratio (O.R) for the independent variable X_i. It gives the relative amount by which the odds of the outcome of interest increases (O.R > 1) or decreases (O.R < 1) when the value of the independent variable is changes by 1 unit. The relative amount by which the odd of the survival outcome increases by a unit change in infant weight is determined by this approach.

Mathematical analysis of the Binary Logistic Model relating Infant Survival Outcomes and Infant weight

We state the model as;

$$Logit\ (p_i) = \alpha_i + \beta_i (x_i - \bar{x}) \quad (10)$$

where,

$$Logit\ (p_i) = \log(\frac{p_i}{1-p_i}), \tag{11}$$

α_i *and* β_i are the logistic regression parameters, p_i is the Infant Survival Rate, the covariate x_i is the average weight of infant in each age group $i = 1,2,3$. In this work, we refer to average weight of infant in each age group simply as infant weight and the overall mean weight of infants as the mean weight of infants for convenience.

From equation (1) we have that;

$$p_i = \frac{1}{1+e^{-[\alpha_i+\beta_i(x_i-\bar{x})]}} \tag{12}$$

Centering the weight of Infants (x_i) at the mean ($\bar{x}$)

We centre the average weight of infants for each age group denoted x_i at the overall mean denoted $\bar{x}$ for convenience. This is because none centering will mean zero weight which is not realistic. Centering at the overall mean weight will help determine the impact the age grouping has on the infant survival rate (p) as shown below.

Centering implies $x_i = \bar{x}$

When this holds, equation (12) becomes,

$$p_i = \frac{1}{1+e^{-\alpha_i}} \tag{13}$$

If $\alpha_i < 0$ then;

$$p_i = \frac{1}{1+e^{\alpha_i}} \tag{14}$$

this has a negative impact on p_i

If $\alpha_i > 0$ then;

$$p_i = \frac{1}{1+e^{-\alpha_i}} \tag{15}$$

this has a positive impact on p_i

Effect of the covariate weight of infant (x_i) on the infant survival rate (p_i)

We establish mathematically, the effect of infant weight x_i on the infant survival rate, p_i of the i^th^

age group. Recall equation (12);

$$p_i = \frac{1}{1+e^{-[\alpha_i+\beta_i(x_i-\bar{x})]}}$$

Observe that, if $\beta_i < 0$ then;

$p_i \to 0$ as $x_i \to \infty$ This shows that increased and decreased magnitude of x_i has the effect of decreasing and increasing p_i respectively. Hence, we term infant weight (x_i) a Non-risk factor of infant survival rate (p_i).

Observe also from equation (2) that if $\beta_i > 0$ then;

$p_i \to 1$ as $x_i \to \infty$. This shows that increased and decreased levels of x_i has the effect of increasing and decreasing p_i respectively. Hence, we term infant weight (x_i) a Risk factor of infant survival rate (p_i).

We model the level of the risk factor (L_{Risk}) as;

$$L_{Risk} = p(\beta_i > 0) * 100\% \tag{16}$$

where $p(\beta_i > 0)$ is the probability of having positive values of β_i. This is the proportion of time (%) that reduced magnitudes of Maternal Age group – Specific Infant Weight have negative effect on infant survival rate.

The Bayesian Binary Logistic Simulation Model relating Maternal Age –Specific Infant Survival Outcomes and Infant weight

As earlier mentioned, the Classical Binary Logistic Modeling Approach can be challenging in the face of incomplete data. On the contrary, the Bayesian Binary Logistic Model does not bow to this challenge (Taeryon et al. 2008). This is because unlike its classical counterpart which considers model parameters as fixed and data as random variables, it considers model parameters as random variables with known probability distributions and data as fixed. Hence it depends chiefly on model parameter sampling and not data sampling.

We therefore develop and implement a Bayesian Statistical Modeling Procedure for modeling the relationship between Maternal – Age Specific Infant Survival Outcomes and Infant Weight. The modeling procedure embeds the Markov Chain Monte Carlo (MCMC) algorithm implemented on an Open Source Software Platform - Windows Bayesian Inference Using the Gibbs Sampler (WINBUG) (Geman and Geman, 1984).

The Bayesian Statistical Simulation Modeling Procedure

Given two faces of the coin; the narcotic drug use prevalence (p_i) for an a maternal age group i and the infant death rate $(1 - p_i)$, we propose the Binomial Likelihood such that;

$y_i \setminus p_i \sim$ Binomial (p_i, n)

Where, y_i is the number of survivals in age group i, the survival rate, $p_i = {y_i}/{n_z}$, n_z is the number of infants delivered by mothers in age group. We state that the computation of p_i per n $(= 1000)$ persons was done in order to determine the observed values of y_i per 1000 persons and for computational ease.

Logistically, p_i is the transformation of the regression mean,

$\alpha_i + \beta_i(x_i - \bar{x})$ and we state that;

$$Logit\ (p_i) = \alpha_i + \beta_i(x_i - \bar{x}).$$

We suppose that the regression parameters α_i and β_i have the priors;

$$\alpha_i \sim Normal(0,0.01),\ \beta_i \sim Normal(0,0.01).$$

In WINBUG syntax, we fit the Bayesian Logistic Regression Model with centered covariate as follows;

Model {

For (i in 1: k) {

$y[i] \sim dbin(p[i],n)$

$logit(p[i] < - alpha[i] + \ beta[i] * (x[i] - mean(x))$

$prob[i] < - step(beta[i] - 0.5)$

$alpha[i] \sim dnormal(0, 0.01)$

$beta[i] \sim dnormal(0, 0.01)$ }

}

Where the data list of number of infant, $y[\ ]$ and infant weight , $x[\ ]$ as well as the initialization list for the model parameter arrays; $alpha\ [\ ]\ and\ beta\ [\ ]$ are defined for each age group i in each zone. k is set as the number of age groups while n is set at 1000. The simulation was run for 100,000 burn-ins after which samples were collected for 300,000 iterations. A thinning of 32 was maintained throughout the simulations and the overlay check box in WINBUG checked to reduce autocorrelation. Other modeling requirements are as stated by the WINBUG Software documentation.

We mention that WINBUG uses the equation;

$$p = \frac{1}{1+e^{-(alpha-beta(x-\bar{x})}} \qquad (17)$$

in computing the Survival Rate (p_i) for each age group, i. This gives the simulated value of p_i.

Model convergence diagnostic check

Model convergence diagnostics was done using history plots, density plots and autocorrelation plots. They plots were produced when the model parameters and measures were monitored on WINBUG. Our approach for investigating convergence issues is by inspecting the mixing and time trends within the chains of individual parameters. The history plots are the most accessible convergence diagnostics and are easy to inspect visually. The history plot of a parameter plots the simulated values for the parameter against the iteration number. The history plot of a well-mixing parameter should traverse the posterior domain rapidly and should have nearly constant mean and variance. The density plots of the model parameters were checked against their actual probability distributions to see whether the right distribution is simulated. This was done for the alpha and beta distribution for each age group i.

Samples simulated using MCMC methods are correlated. The smaller the correlation, the more efficient the sampling process. Though, the Gibbs, MCMC algorithm typically generates less-correlated draws, there is a need to monitor the autocorrelation of each parameter to ensure samples are independent. The autocorrelation plot that comes from a well-mixing chain becomes negligible fairly quickly, after a few lags. This was achieved for the model parameters and measures.

RESULTS AND DISCUSSION

The study results include some results of the Classical Binary Logistic Modeling approach such as; the fitted Binary Logistic Model, the Hosmer – Lemeshow goodness of fit test and the model outcome classification table. The study results further include the actual infant survival rate and the infant weight for each maternal age group in the population of study. Before the results of the Bayesian Statistical Simulation Modeling Approach, we present the Chi-square test of homogeneity of number of Infant Survival Outcomes across maternal age groups. Results of the Bayesian Statistical Simulation include the simulation model parameter and measure values for each maternal age group, results of the model diagnostic checks for the first maternal age group; these include history plots, density and autocorrelation plots. Further result include the distribution of Maternal Age group - Specific impact of infant weight on infant Survival Outcome, a distribution of infant weight factor status and an infant weight risk magnitude across the maternal age groups.

Table 1 : Classical Binary Regression Model Parameters

	Model coefficient (B)	S.E.of B	Value of Wald Statistic	Degree of Freedom (df)	P _ value	Exp(B)	95% Confidence .Interval for EXP(B)	
							Lower	Upper
ChildWT	.856	.190	20.392	1	.000	2.354	1.623	3.413
Constant	.305	.567	.289	1	.591	1.356		

Model: Logit(p) = 0.305 + 0.856 * Infant weight.

The model shows that infant weight was retained as the only covariate (table 1) and since mother age is not retained, the model cannot be Maternal Age-Specific which is the modeler's interest. This as earlier mentioned led to the chi - square test of homogeneity of the distribution of infant survival outcomes across three classes of maternal age group at child birth.

Table 2 : Contingency Table for Hosmer and Lemeshow Test

	State of Infant Birth (dead) = 0		State of Infant Birth (alive) = 1		Total
	Observed	Expected	Observed	Expected	
1	12	11.562	75	75.438	87
2	3	8.049	116	110.951	119
3	6	2.726	41	44.274	47
4	4	5.403	97	95.597	101
5	4	6.372	132	129.628	136
6	9	5.044	116	119.956	125
7	6	4.799	129	130.201	135
8	2	2.963	92	91.037	94
9	4	3.081	126	126.919	130

Chi-square value = 13.049, degree of freedom (df) = 7, p_value = 0.071

Table 3 : Model outcome classification table

Observed		Predicted		
		State of Infant Birth		Percentage Correct
		0	1	
State of Infant Birth	0	0	50	0.00
	1	0	924	100.0
Overall Percentage				94.9

The cut value = 0.50

The fitted Binary Logistic Model is well validated with a good Hosmer – Lemeshow goodness of fit test result (p > 0.05) and overall percentage model correct classification of 95% survival outcomes, as shown in table 2 and table 3.

Table 4: Distribution of actual infant survival rate and weight across maternal age groups

Age group (years)	Actual survival rate	Average weight of infant (kg)
15-25	0.900	3.261
26-35	0.869	3.150
<= 36	0.887	3.687

The actual infant survival rate and the average infant weight for each maternal age group are captured on table 4. These values were computed from actual aggregate data on the number of infant who survived and actual weights of infants. The number of infants who survived per 1000 persons were determined for each maternal age-group and used as number of trials for the Binomial distribution in the course of the Bayesian statistical modeling. The average weights of infant data sets for each age group also serve as input to the model.

Table 5 : Frequency of survivals

	Observed frequency	Expected frequency	Residual
15 - 25 years	273	197.0	76.0
26 - 35 years	232	197.0	35.0
Greater than or equal to 36 years	86	197.0	-111.0
Total	591		

Chi-square value = 98.081, Degree of freedom = 2, p_value = 0.00

Table 5 shows the distribution of survival across maternal age groups. The test shows that the distribution of infants who survived differ significantly across the age groups ($P < 0.05$) hence the need for Maternal Age-Specific Infant Survival Outcome Prediction Model. The three age groups identified were: 15 – 25, 26 – 35 and greater than or equal to 36 years.

Table 6: Model parameters and measure values for each maternal age group

Node	Mean	Sd	MC error	2.50%	median	97.50%	start	sample
alpha[1]	2.183	1.048	0.004257	0.1187	2.178	4.247	100000	50001
alpha[2]	1.774	2.1	0.01352	-2.327	1.763	5.917	100000	50001
alpha[3]	1.859	3.044	0.02579	-4.083	1.855	7.842	100000	50001
beta[1]	-0.2818	9.972	0.04011	-19.9	-0.2977	19.39	100000	50001
beta[2]	-0.5623	9.707	0.06228	-19.5	-0.6125	18.57	100000	50001
beta[3]	0.6415	9.491	0.0804	-18.01	0.6503	19.17	100000	50001
p[1]	0.9009	0.009422	4.12E-05	0.8817	0.9012	0.9185	100000	50001
p[2]	0.869	0.01063	4.94E-05	0.8475	0.8692	0.8891	100000	50001
p[3]	0.887	0.01004	4.47E-05	0.8666	0.8873	0.906	100000	50001
prob[1]	0.4671	0.4989	0.002073	0	0	1	100000	50001
prob[2]	0.4558	0.498	0.002964	0	0	1	100000	50001
prob[3]	0.506	0.5	0.003665	0	1	1	100000	50001

Note (i) The numbers 1 – 3 indicate the maternal age groups

(ii) model parameters are alpha and beta

(iii) model measures are infant survival rate (p) and risk magnitude of infant weight (prob)

Model parameters and measure values for each maternal age group are captured on table 6. Details on these tables include the mean, standard deviation, Monte Carlo Simulation Error, median and 95 % credible interval. Estimates of the model parameters (alpha and beta) and their measures infant survival rate (p) and magnitude of risk (prob (beta > 0)) are their respective mean values. The values of alpha and beta for each maternal age group, helps to determine the relationship between the infant survival outcomes and infant weight. This is achieved when they are plugged into equation (1). Observe the close values of the actual infant survival rate (p) on table 4 for each maternal age group and the simulated values $p[i])$ on tables 6. This also validates the model.

Table 7: Distribution of model measures across maternal age groups

Age group	Sign of model parameter (α)	Age group impact on infant survival rate	Sign of model parameter (β)	Infant weight factor status	Risk magnitude of infant weight (%)
15 – 25 years	+	Positive	-	Non-risk factor	-
26 – 35 years	+	Positive	-	Non-risk factor	-
>= 36 years	+	Positive	+	Risk factor	50.60

As established in the mathematical analysis, the sign of the model parameter alpha assists in determining the impact of an age group (positive or negative) on infant survival rate while the sign of beta assists in determining whether the infant weight is a risk factor of infant survival rate or not. The level of this risk is computed using equation (6). As earlier mentioned, it is the proportion of time (%) that reduced magnitude of infant weight of a maternal age group has negative effect on the infant survival rate. The overall results shows that all the three maternal age groups have positive impact on infant survival rate while, the weight of infants delivered by mothers in the 15 – 25 years and 26 – 35 age groups are none risk factors to infant survival rates but those of infants delivered by mothers in the 36 years and above age group pose as risk factor to infant survival rate. See table 7 for details. The Binary Logistic Model has become richer as it has become Maternal Age – Specific. Hence, once mother age is known, an appropriate model can be chosen to predict infant survival outcome or rate given infant weight.

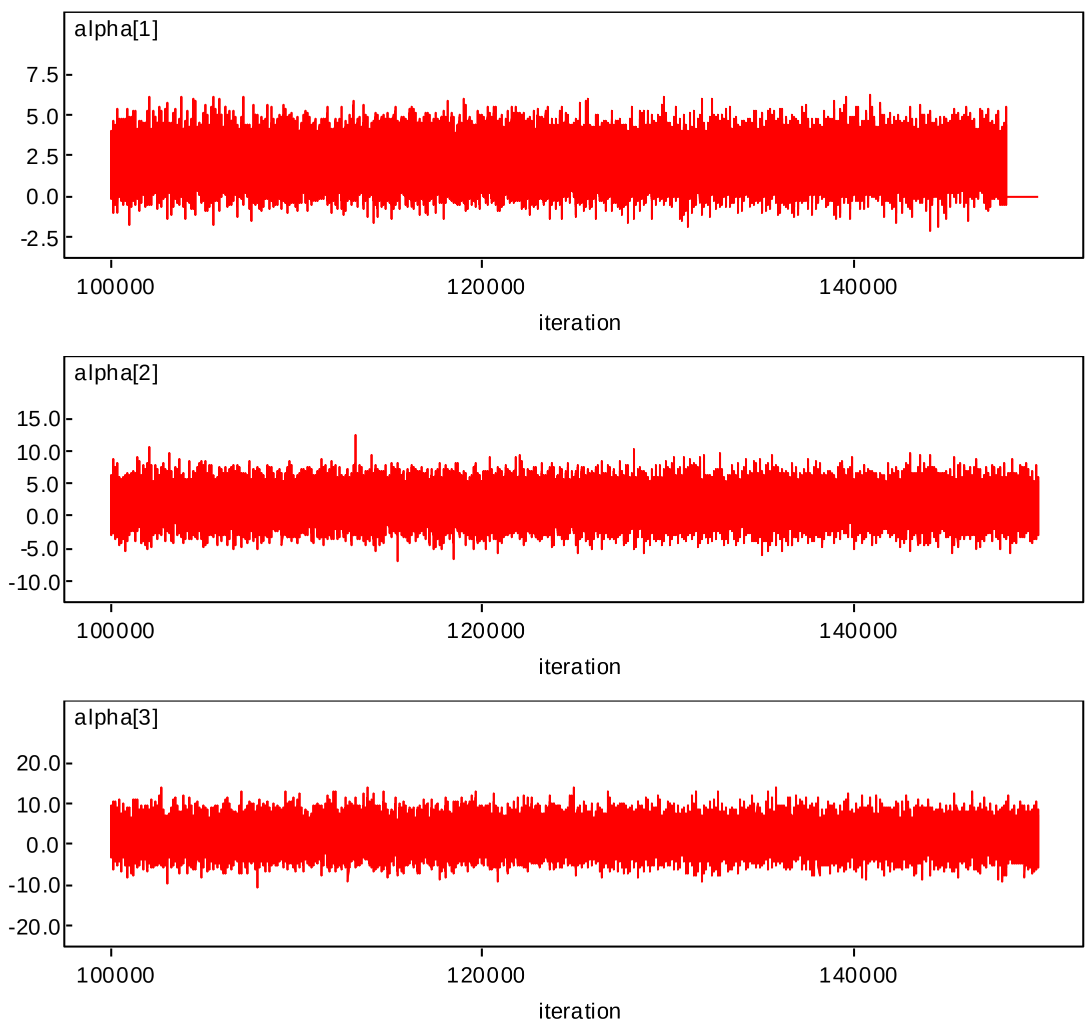


Fig. 1: History plots of model parameter alpha

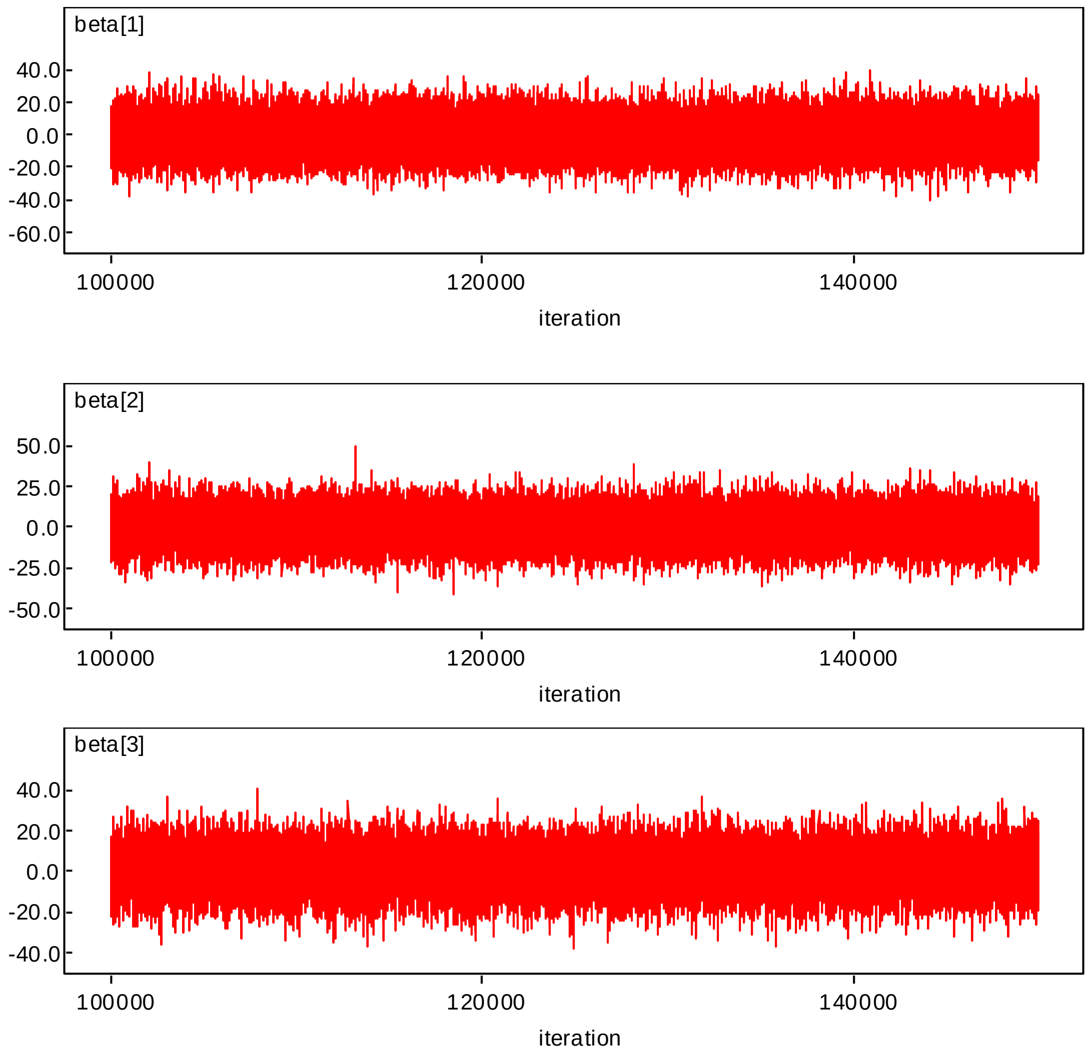


Fig. 2: History plots f model parameter beta

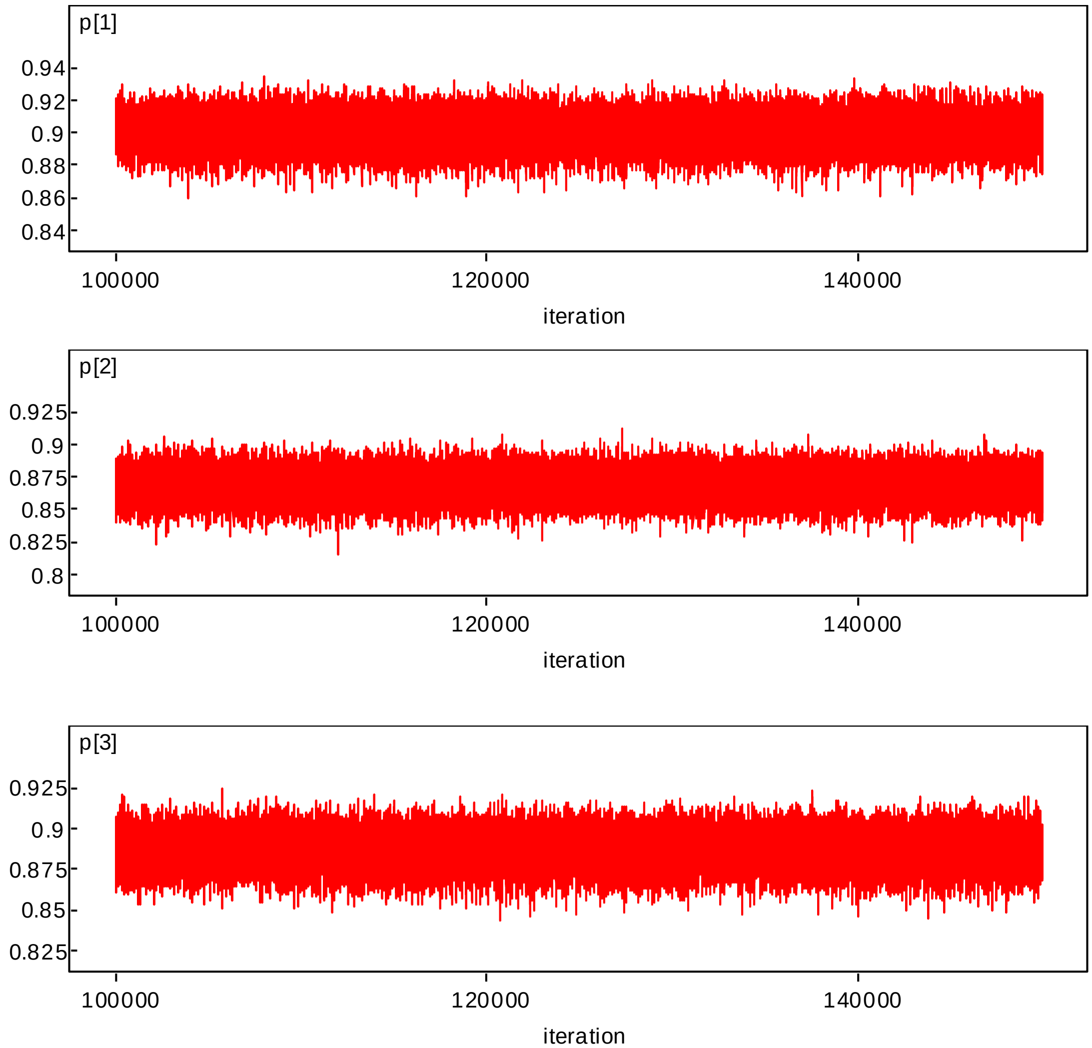


Fig. 3: History plots of infant survival rate (p)

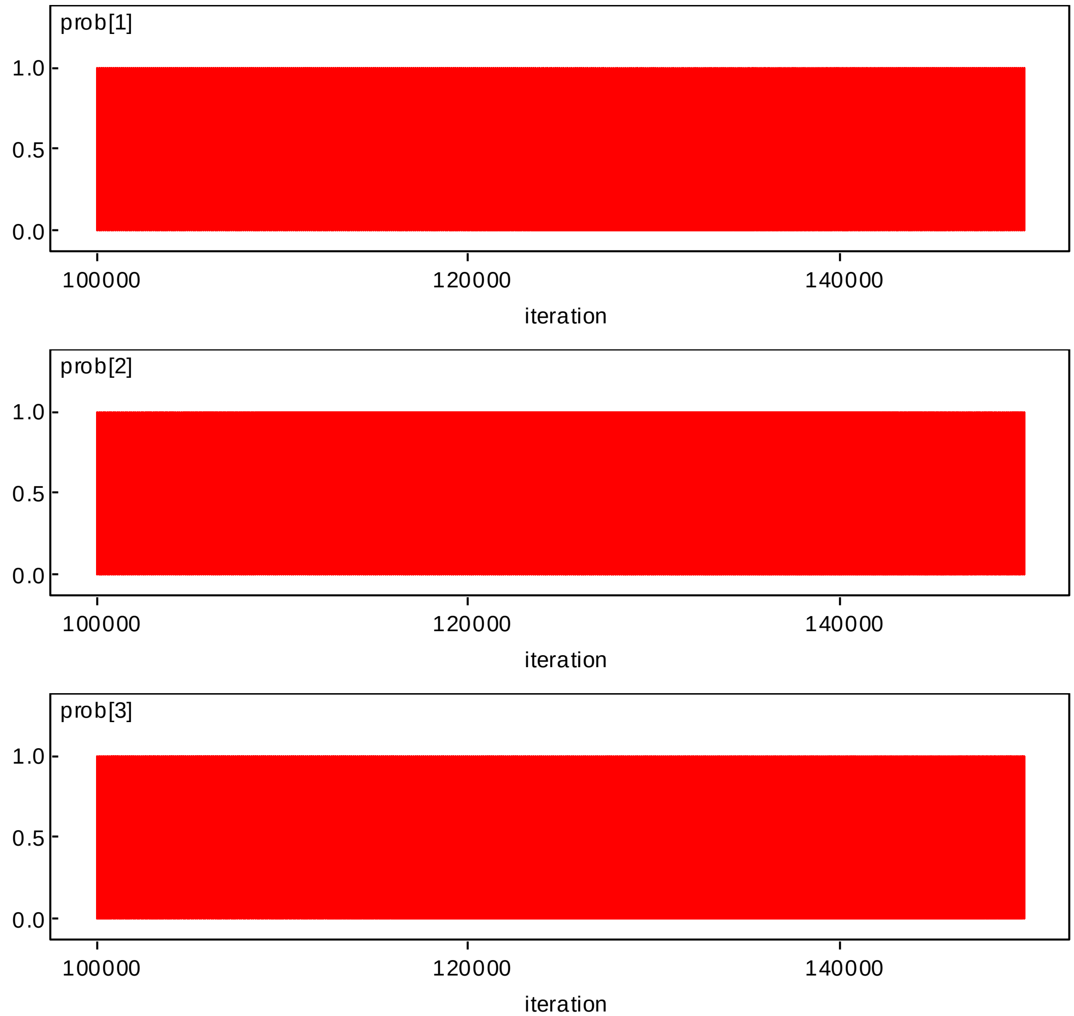


Fig. 4: History plots of risk magnitude of infant weight

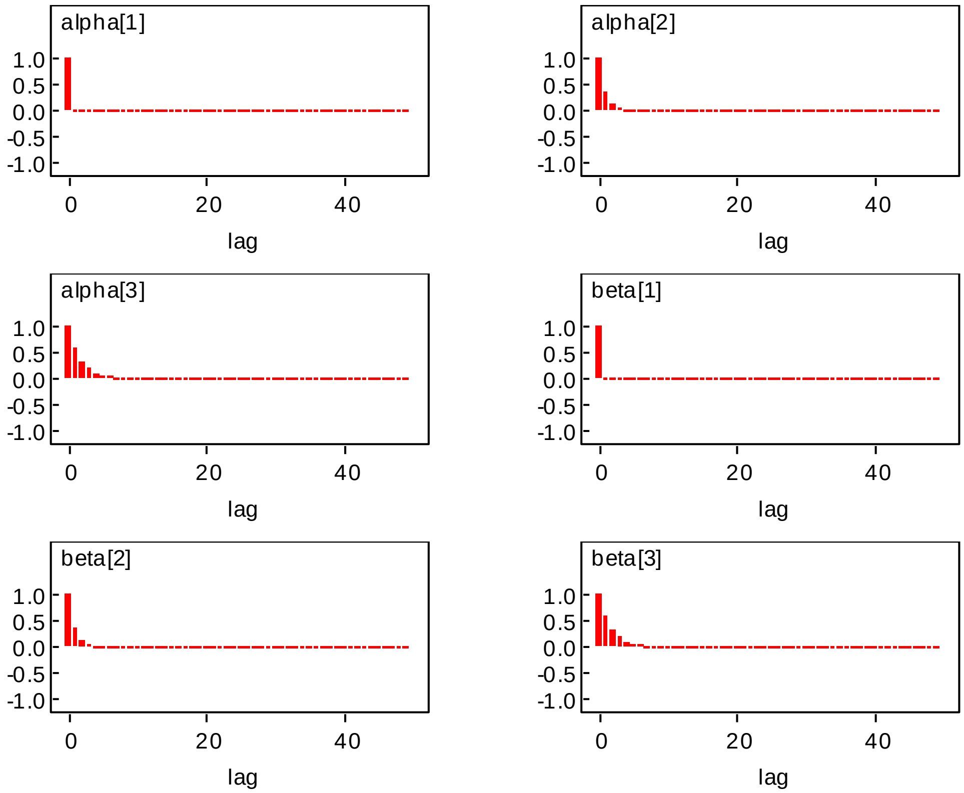


Fig. 5: Autocorrelation plots of model parameters alpha and beta

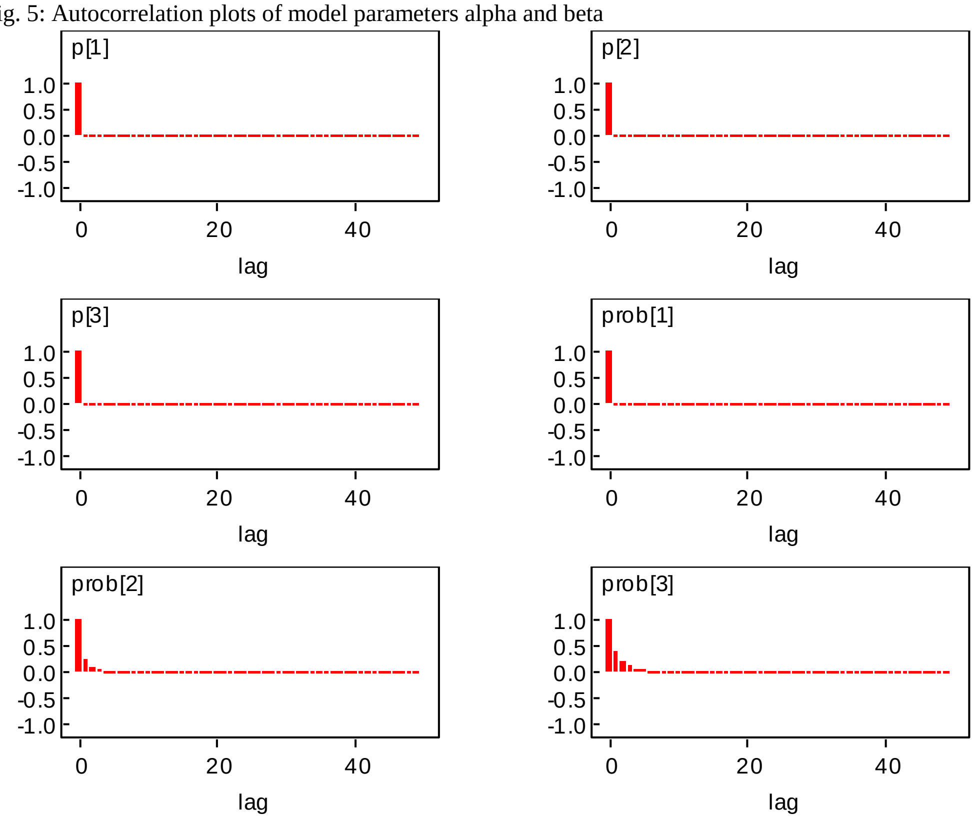


Fig. 6: Autocorrelation plots of infant survival rate (p) and risk magnitude of infant weight (prob)

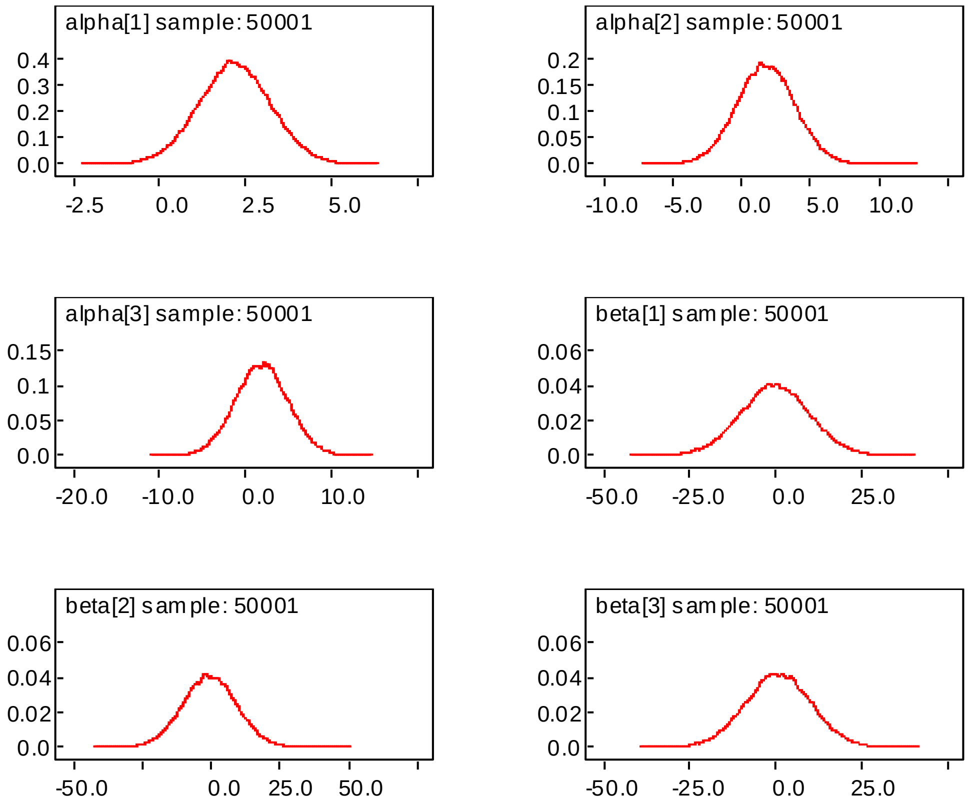
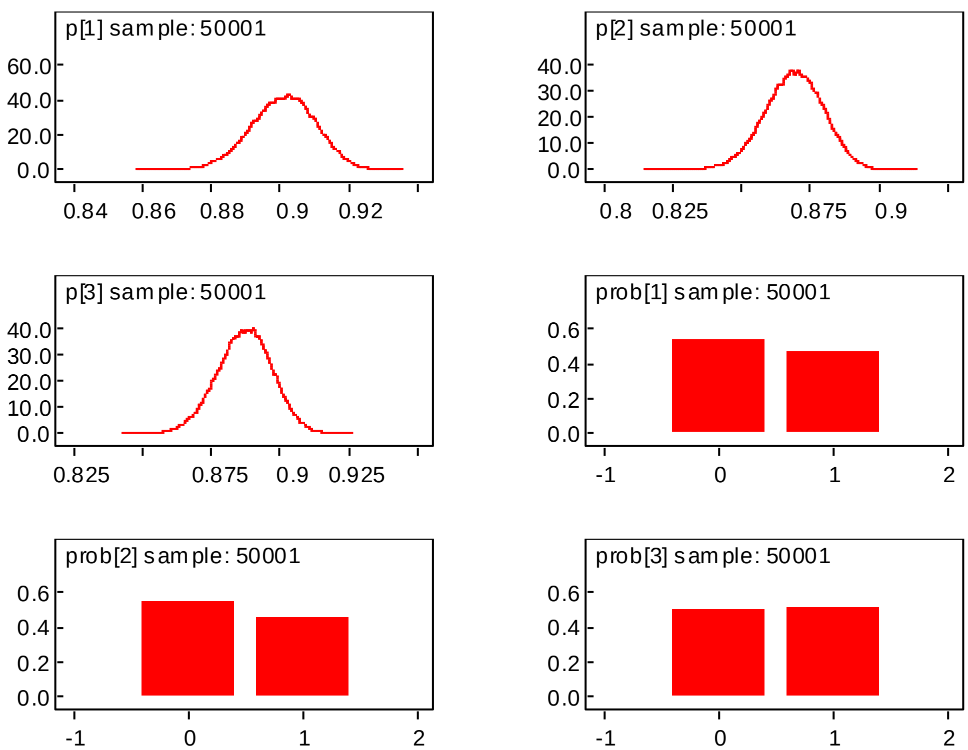


Fig. 7: Density plots of model parameters alpha and beta



Fig, 8: Density plots of infant survival rate (p) and risk magnitude of infant weight (prob)

As earlier mentioned, the incompleteness of this data limits the use of the Classical Logistic Regression Model (Taeryon et al., 2008). Tripathi et al (2019) further confirm this when they consider a Logit model for assessing child mortality under three maternal parity (3, 4 and 5) using a Bayesian swatch. Parity according to them is the number of child birth experienced by the infant's mother. The grouping of the original data to reflect the parity definitely reduce the sample size, hence the use of Classical Bayesian Approach becomes futile. This limitation calls for the development of a Bayesian Statistical Modeling Procedure using the MCMC Gibbs algorithm on the WINBUG platform.

After the development of the model, convergence diagnostic checks were conducted for each model parameter and measure in order to ascertain model adequacy. The history plot, density plots and autocorrelation plots were used for this purpose. See figures 1 – 8. Observe that the history plots shows that the model parameters and measures are well – mixed. This is because they traverse the posterior domain rapidly with nearly constant mean and variance. The model prior distribution for alpha and bêta is normal (0, 0.01). The density plots of these priors reflect this distribution which further validates the model. The autocorrelation plots of each parameter and measure depict the independence of the samples generated. This is because the autocorrelations become negligible fairly quickly, after a few lags.

CONCLUSION AND RECOMMENDATION

A Bayesian Binary Logistic modeling procedure was developed for predicting Maternal Age Specific Infant Survival Outcome using incomplete data. The three maternal age groups; 15 - 25 years, 26 - 35 years and 36 years and above have positive impact on infant survival rate. The weight of infants delivered by mothers in the 15 - 25 years and 26 - 35 years age groups are none risk factors to infant survival rates but those of infants delivered by mothers in the 36 years and above age group pose as risk factor to infant survival rat

The study recommends that this modeling procedure should be applied to similar modeling problems with the challenge of incomplete data and that the models be used to compliment doctors efforts in antenatal care delivery.

REFERENCES

Gemperil, A. (2004). Spatial Patterns of Infant Mortality in Mali. The effect of Malaria Endemicity. *American Journal of Epidemiology*, 159: 64-72

German, S and German, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restorationof images. IEE Trans. Pattern Anal Mach Intell. 6:721-741.

Koissi, M . C. and Hogens, G. (2005). Using WINBUGS to study Family Frailty in Child Mortality with Application to Child Survival in Ivory Coast. African Population Studies, 20: 1-27.

Long, J.S. (1997). Regression Models for Categorical Data and limited dependent variables. Thousank Oaks, CA: Sage Publications.

Peduzzi, P., Concato, J., Kemper, E. , Holford, T.R and Fesintein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*. pp. 1373-1379.

Selemani, M., Mwanyangala, M.A., Mrema, S. (2014). The effect of mother's age and other related factors on neonatal survival associated with first and second birth in rural, Tanzania: evidence from Ifakara health and demographic surveillance system in rural Tanzania. BMC Pregnancy Childbirth 14(240) doi:10.1186/1471-2393-14-240

Taeryon, C., Mark, J. S., Ketra, A. S. and Mitchell, J.S. (2008). A Bayesian approach to a Logistic Regression Model with Incomplete data'', *Biometrics.* 64: 424 - 430.t

Tripathi, A., Singh, G. and Singh, A.(2019). Model for Assessment of Child Mortality under Different Parity : A Bayesian Swatch. Journal of Biosciences and Medicines,7: 113-126. Doi: 10.4236/jbm2019.75014