

Bayesian Modelling of Tuberculosis Risk Factors in South Africa 2014

Hilda Dhlakama^{1,*} and Siaka Lougue²

¹Department of Statistics, University of Johannesburg, P.O Box 524, Auckland Park, Johannesburg, 2006, South Africa

²School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X54001, Durban, 4000, South Africa

Abstract: *Background:* Although the number of deaths has declined since 2007, Tuberculosis (TB) continues to be the number one cause of death in South Africa. To create a country free of TB, there is need for continued research to explore models that will provide the Department of Health with new interventions.

Aim: This study was aimed at identifying the risk factors of active self-reported TB prevalence for South Africa in 2014.

Methods: The Frequentist Logistic Regression (FLR) approach was applied on a sample of 19213 individuals taken from the National Income Dynamics Survey (NIDS) wave data. Bayesian analysis with non-informative priors were used to model Wave 1 to 3 data and elicitation of the obtained posterior density parameters by averaging done to obtain the informative priors used to model wave 4. The wave 4 results obtained under the two estimation approaches were compared as well as the results for non-informative and informative priors.

Results: The findings show that self-reported TB prevalence is higher than the reported 1%, Human Immuno Deficiency Virus (HIV) remains a major threat to TB and Eastern Cape is the province mostly affected by TB with Limpopo recording the least prevalence. Poor living conditions and lower socio-economic conditions continue to be drivers of TB whilst English illiteracy, lack of Secondary/Tertiary education, alcohol consumption, marital status, gender and age groups also influence TB progression to disease. FLR yielded similar results to Bayesian with non-informative priors whilst the results are more precise for informative priors.

Conclusion: This study identified individuals and communities at risk of developing active TB disease in South Africa.

Keywords: Tuberculosis, HIV, South Africa, National Income Dynamics Survey, Bayesian analysis, Frequentist Logistic Regression.

INTRODUCTION

World Health Organisation (WHO) reported that approximately one-third of the World's population is infected with TB, most of which occur in the low-income countries [1]. In 2013 TB accounted for 8.8% of all deaths in South Africa. The South African department of health also reported that 73% of TB patients are HIV positive [2]. The recently launched WHO post-2015 Global Strategy aims at "ending the global TB epidemic by reducing deaths by 95% and cutting new cases by 90% between 2015 and 2035 and to ensure that no family is burdened with catastrophic expenses due to TB" [3].

WHO also reported 450 000 cases of active TB which means that approximately 1% of the South African population develop active TB every year, placing South Africa on third position of highest incidence after India and China [4]. The Department of Health in South Africa launched the Stop TB strategy

under the National TB Control Programme National Strategic Plan (NTCP) whose objectives are to reduce TB by effective treatment, community awareness and research [5]. In bid to reduce TB infections and deaths, many research on TB have been conducted over the years and some are underway but it is still a WHO and the Department of Health (SA) goal to further reduce TB in South Africa and globally. This study was aimed at understanding adult (15+ years old) self-reported TB by developing models that best describe individual and household risk factors in South Africa, in 2014.

As a disease of poverty, the risk factors of TB are mostly associated with poor living conditions, overcrowding, substandard living or working conditions. [6] investigated the interplay of multiple factors affecting the prevalence of TB in a low income township in Grahamstown, Eastern Cape Province of South Africa called Rhini. Overcrowding and roof leakage were the main contributors towards the probability of a household experiencing TB whereas higher social capital significantly reduced this probability. Using a hierarchical regression model, [7] evaluated individual, household and community level risk factors for self-reported TB disease and reached

*Address correspondence to this author at the Department of Statistics, University of Johannesburg, P.O Box 524, Auckland Park, Johannesburg, 2006, South Africa; Tel: +27115596021; Fax: +27115594896; E-mail: hildad@uj.ac.za

the same conclusions as [6]. They used data from South Africa Demographic and Health Survey and the national census 1996. The study also found out that having low education, being unemployed, living in a household with low level of wealth and living in a community with high levels of income inequality are independently associated with an increased risk of having ever been diagnosed with TB. Many other publications attest to TB being a disease of poverty among them [8,9].

The most frequently used method to estimate parameters for statistical models are known as the frequentist and Bayesian approaches. The frequentist approach assume that unknown population parameters are fixed constants, and they interpret probability as long-run relative frequencies. Inference is based on the behaviour of a sample statistic while hypothesis testing relies on the behaviour of a test statistic under the null hypothesis [10]. According to [11], "the frequentist statistical procedures are judged by how well they perform over an infinite number of hypothetical repetitions of the experiment". The frequentist approach is therefore concerned with parameter estimation and provide statistical inference based on the classical P-value, the level of significance, the power and confidence interval [10]. On the contrary, the Bayesian approach treats parameters as random variables, the rules of probability are used to make inference about parameters. "The parameter is stochastic and inference is done given the observed data" [10]. The data is used to update prior belief regarding the parameters [12]. Bayesian inference cannot be done without a prior distribution which is the distinguishing feature of Bayesian methodology. A prior distribution of a parameter is the probability distribution that represents your uncertainty about the parameter before applying the current data and it must be subjective. This approach uses the posterior distribution which comes from two sources: the prior distribution and the observed data. These two approaches only take different routes to achieve the same goal, which is: to discover the true parameter value of the parameter.

MATERIALS AND METHODS

The data used for this research is Wave data from South African NIDS. It is the first panel study of a nationally representative sample of 28,000 individuals in 7 300 households across the country that has been tracking their lives since 2008. The data is freely available from the NIDS webpage. It contains data on

income, health, education, migrations, wealth, labour and more. The survey continues to be repeated with these same household members every two years. NIDS examines the livelihoods of individuals and households over time. The target population for NIDS were private households and residents in worker's hostels, convents and monasteries. The ethics approval was granted by the University of Capetown Commerce Faculty Ethics Committee.

A detailed description of this data is given elsewhere [13]. To sample the households that were included in the base wave, a stratified, two-stage cluster sampling design was used. The survey sample was selected using the Stats SA's Master Sample of more than 3000 Primary Sampling Units. Four hundred PSUs were proportionally selected from the 53 District Councils (DCs) for use in the NIDS survey. Stratified sampling was used for this selection. Within each PSU, 8 samples of dwelling units were selected from the Master Sample using systematic sampling. 2 of them were allocated to NIDS these are referred to as "clusters". All households residing at the selected dwelling unit and all household members automatically becomes NIDS sample members.

Variables

For this study, the focus is on self-reported TB prevalence in South Africa for Wave 4 (2014). The respondents were asked if they were diagnosed with TB, if they were on medication or if they were not on medication but still having TB. The dependent variable for this study includes patients who reported to be on TB medication and those that were not on medication but still had the disease. This variable was coded 0-if the responded was not on TB medication, 1-for the presence of TB or was not on TB medication but still had TB. The independent variables were at individual and household levels and were obtained from past research and simple stepwise logistic regression. All variables were categorical. A sample of size 19213 with complete cases only was used.

Methods

Individual and household risk factors of active TB were analysed using the logistic regression modelling. Combined stepwise regression was used to select the best model, one that best explains the data, the one with the lowest Akaike Information Criterion (AIC) [14]. The significant variables obtained from the best model were then used for further analysis. The frequentist and

Bayesian approaches were both used for inference. Non-informative priors were used to model Wave 1 to 3 data considering the complete cases only. Elicitation of prior knowledge from historical data was done in order to obtain informative priors to use for Wave 4. This was done by averaging the individual posterior density parameters obtained for Wave 1 to 3. The results for frequentist and Bayesian approaches were compared.

The main purpose of logistic regression is to give an overview of the relationships and strengths between the dependent and independent variables. A special case of Generalised Linear Models (GLMs), the logistic model is used for binomially distributed data and is used to represent how a binary (dichotomous) outcome variable depends (on or is related to) a set of independent variables. Most studies on risk factors of diseases use the FLR models under the frequentist method of estimation [15,16]. The logistic regression function is used because the relationship between the two variables, dependent and independent, is not a linear function. The multiple linear logistic regression model with covariates x_1, x_2, \dots, x_k asserts that the probability p of occurrence of a binary event of interest may be represented as [17];

$$\text{Logit } [p] = \log \left[\frac{p}{1-p} \right] = \beta_0 + \sum_{j=1}^k \beta_j x_j$$

where; β_0 is the intercept, $\beta_1, \beta_2, \dots, \beta_k$ are covariate coefficients and $\frac{p}{1-p}$ is the odds ratio.

In the frequentist inference approach, model selection is done using either forward, backward or combined stepwise regression. Combined stepwise regression was used for is a combination of forward stepwise and backward elimination. Variables are added or removed early in the process. The Maximum Likelihood Estimation (MLE) is used for estimation. The reliability of the model estimation also depends on the logistic model being appropriate which implies the need for checking the model adequacy.

The Bayesian approach offers a number of advantages over the frequentist approach and some of them by [11] are listed below:

1. It has a unified framework, a single tool, the Bayes' theorem which is useful to all situations opposed to the frequentist approach which has different tools. Namely random effects, hierarchical, missing variables, nested or non-nested models.

2. The Bayes' theorem gives an easier way to find the predictive distribution for future values than the frequentist way.
3. Bayesian statistics uses both sources the prior information we have about the process and the information contained in the process.
4. Bayesian statistics often lead to more accurate models in terms of predictive performance. They use credible intervals that give the direct probability statement that a parameter is contained in a certain interval whereas the frequentist uses confidence intervals which are interpreted as probability intervals. A 95% credible interval is much narrower than the 95% confidence interval.
5. Quantifies all aspects of uncertainty through probability. Probability is the central tool. [18].

In Bayesian models, the likelihood of the observed data y given a set of parameters $\theta = (\theta_1, \dots, \theta_d)$ denoted by $P(y|\theta)$ is used to modify the prior beliefs $P(\theta)$ [19]. The updated knowledge based on the observed data and the information contained in the prior density follows from standard probability relations:

$$P(y, \theta) = P(y|\theta)P(\theta) = P(\theta|y)P(y)$$

Therefore, the posterior density can be written as:

$$P(\theta|y) = P(y|\theta)P(\theta)$$

The common fundamental algorithm in Bayesian analysis is the Markov Chain Monte Carlo (MCMC). The two most important MCMC procedures are the Gibbs sampler and the Metropolis (Hastings) algorithm [10]. In this study the Gibbs sampler procedure was used. Popularised by [20], the MCMC simulates a discrete time Markov chain on the state space of θ whose steady state distribution is the posterior distribution of interest $P(\theta|y)$. The Gibbs sampler generates dependent sequence (a chain) of random variables, initialised at the starting value $\theta^{(0)}$. It approximates the true posterior density $P(\theta|y)$ using a chain of samples drawn from the density. The collection of M samples, $\{\theta^{(1)}, \dots, \theta^{(M)}\}$ where a sample is assumed to be drawn from the posterior. Specifies that the distribution of :

$$\theta^{(i)}, \theta^{(i+1)} \text{ is independent from } \theta^{(i-1)}, \theta^{(i-2)} \dots$$

Therefore, in a probabilistic notation, $P(\theta^{(i+1)} | \theta^{(i)}, \theta^{(i-1)}, \dots, y) = P(\theta^{(i+1)} | \theta^{(i)}, y)$.

An initial portion of the chain has to be discarded, this is called the burn-in part [10].

The convergence diagnostic used was Brooks-Gelman-Rubin (BGR) diagnostic. It values the variance of the values sampled by each of the three chains. At convergence, the BGR diagnostic plots shows a ratio of 1.

RESULTS AND DISCUSSION

The TB prevalence rates are presented in this subsection regardless of whether the χ^2 test is significant or not.

TB prevalence for this sample was 1.6%. TB is more prevalent in other races combined, 1.8%, compared to the African race (1.6%). TB prevalence is higher in males 1.8% than females. On marital status, the divorced/widowed has the highest prevalence of 2.3% whereas the married/living together and the never married both have a prevalence of 1.5%. People with secondary/tertiary education as their highest level have the least TB prevalence, 1.1% whereas those with no form of education and primary have a prevalence of 3.0 and 3.2 respectively. The IsiXhosa speaking people have the highest TB prevalence of 1.9% followed by those speaking other languages combined, with the isiZulu speaking having a prevalence 1.5%. The unemployed have a higher TB prevalence of 1.9% whereas those employed have a prevalence of 1.3%. The age group most infected by TB is the 45-59 years, 2.9% followed by the 30-44 years old at 2.3%. The least TB infected age group is the 15-29 years, 0.78%. TB prevalence is higher in those that consulted about their health in the last two years, 2.0% than those who have not consulted about their health in more than two years. Those that do not exercise have a higher prevalence of TB, 1.8% than those who exercise, 1.3%. People who value religion have 1.7% prevalence of TB compared to 1.5% of those who do not see religion as important. Those with no religion or other religions than Christianity have a higher prevalence rate of 2.1%. The people who smoke or drink alcohol have a higher TB prevalence. 11.5% of people who are HIV positive have active TB compared to 2.5% who are either negative or did not disclose their status. On the issue of illiteracy, people who cannot read either their home language or English have the same TB prevalence of 3.4% where those who can read home language and English have a prevalence of 1.5 and 1.3 respectively. Those who perceived their health as not good had a TB prevalence of 5.8% whereas those who perceive their health to be good have a prevalence of 1.1%.

A look at the household variables. Households with taped water have less TB prevalence compared to households with other sources of water. 1.5% and 2.4% respectively. TB prevalence for households with flushing toilets is 1.5% whereas those with no flushing toilets has a higher prevalence of 1.8%. Households with below average household income and are below the food poverty line have higher TB prevalence, 2.1% and 1.7% respectively compared to households above average and above food poverty line, 1.3% and 1.5% respectively. There is not much difference in TB prevalence between households with shared or unshared toilets. Households using electricity for cooking or lighting have the same TB prevalence of 1.6% which is less than houses using other sources of energy. Overcrowded households have a prevalence of 1.8% which is higher than the 1.6% for non-overcrowded households. TB is more prevalent in Informal settlements (2.8%) compared to the 1.5% for formal/ traditional dwelling types. Households receiving some form of social grant have a higher TB prevalence of 1.7% as opposed to households that do not receive any social grant (1.5%). The descending order of TB prevalence in provinces is Eastern Cape, Northern Cape, Western Cape, KwaZulu Natal, Mpumalanga, North west, Gauteng, Limpopo with 2.8%, 2.3%, 1.9%, 1.6%, 1.5%, 1.4%, 0.9% and 0.7% respectively.

The R version 3.2.2 was used for the classical approach analysis and the best model selected by stepwise regression yielded the results in Table 1.

The variables that best describe this data on prevalence of self-reported TB in order are: perceived health status, self-reported HIV status, Consultation about health, Province of residence, ability to read English, highest education level, household dwell type, household income level, marital status, alcohol, household source of income and employment status. Individuals who perceive their health status as not good have a log odds of 3.45, 95% C.I (2.65: 4.47). This means that they are 245% more likely to develop active TB than those who perceive their health to be good. This could be because since they will be TB patients, their health status will be compromised. Self-reported HIV negative individuals or those that did not declare their status 82% less likely to develop active TB compared to their HIV positive counterparts. The log odds was 0.18. The log odds of people who last consulted about their health more than two years is 0.26. This means they are 74% less likely to develop TB.

Table 1: Parameter Estimates of the Best Stepwise FLR

	Variables	Estimate	Std. Error	z value	Pr (> z)	OR: Odds Ratio	2.5%	97.5%
(Intercept)	Intercept	-1.94	0.30	-6.44	0.00	0.14	0.08	0.26
(hlsta)2	Health status not good	1.24	0.13	9.28	< 0001	3.45	2.65	4.47
(hiv)2	HIV Negative	-1.73	0.19	-9.24	<0.001	0.18	0.12	0.26
(con)2	Last consult >2years	-1.35	0.25	-5.41	<0.001	0.26	0.15	0.41
(prov)2	Eastern Cape	0.15	0.21	0.71	0.48	1.16	0.77	1.78
(prov)3	Nothern Cape	0.04	0.24	0.16	0.87	1.04	0.64	1.67
(prov)4	Free State	-0.27	0.28	-0.98	0.33	0.76	0.43	1.30
(prov)5	KwaZulu Natal	-0.42	0.21	-2.01	0.04	0.66	0.44	1.00
(prov)6	North West	-0.46	0.30	-1.53	0.13	0.63	0.34	1.12
(prov)7	Gauteng	-0.62	0.27	-2.27	0.02	0.54	0.31	0.91
(prov)8	Mpumalanga	-0.26	0.27	-0.94	0.35	0.77	0.44	1.31
(prov)9	Limpopo	-1.01	0.33	-3.10	0.00	0.36	0.19	0.67
(readen)2	Read English_No	0.39	0.18	2.22	0.03	1.48	1.04	2.09
(gender)2	Gender_Female	-0.43	0.13	-3.41	0.00	0.65	0.51	0.83
(edu)2	Education_Secondary/Tertiary	-0.57	0.15	-3.66	0.00	0.57	0.42	0.77
(edu)3	Education_None	-0.43	0.20	-2.10	0.04	0.65	0.43	0.97
(hhdwtyp)2	Dwelling type_Informal	0.44	0.16	2.66	0.01	1.55	1.11	2.13
(hhincome)2	Household income_below average	0.27	0.12	2.25	0.02	1.31	1.04	1.66
(marstt)2	Marital status_Divorced/ widow	0.14	0.16	0.83	0.41	1.15	0.83	1.58
(marstt)3	Marital status_Never married	0.37	0.14	2.58	0.01	1.45	1.10	1.92
(alc)2	Alcohol_non drinker	-0.27	0.13	-2.07	0.04	0.76	0.59	0.99

Of the nine provinces, using Western Cape as the baseline, only three showed a significant difference. These are KZN, Gauteng and Limpopo with log odds of 0.66, 0.54 and 0.36 respectively. An individual is less likely to develop active TB by 34%, 46% and 64% respectively. Ability to read English is positively related to TB with a log odds of 1.48 95% C.I (1.04: 2.09). This means that people who cannot read English are 48% more likely to get infected by TB compared to people who can read the language. This result is similar to the level of education. People with Secondary or tertiary education as their highest qualifications are negatively associated with active TB with a log odds of 0.57, 95% C.I (0.42- 0.77). They are therefore 43% less likely to get active TB than those with primary education. The log odds for those with no form of education whatsoever is 0.65, 95% C.I of (0.43-0.97). This means they are 35% less likely to get TB infected than those with primary education. Odds of individuals developing active TB was negatively associated with gender with a coefficient of -0.44, log odds of 0.65, 95% C.I (0.51-0.83). The odds of females of getting active TB is 35%

less than the odds for males. The odds of active TB was positively associated with marital status. The widowed/divorced are not significantly different from the married as far as developing active TB whereas the never married are significant with the log odds of 1.45 are 45% more likely to have active TB than the married people. People who drink alcohol are 24% less likely to get infected by TB than the non-drinkers. At household level, only dwelling type and household income level are significant. There is a positive association between TB prevalence in households living in Informal dwellings. The logs odds is 1.55 (95% C.I: 1.11- 2.13). The odds of people living in informal dwellings are 55% more than those living type of formal/traditional dwellings. Lastly households with income below average are 31% more likely to have members infected with TB than households that have average or are above average income.

The informative priors used are shown in Table 2.

OpenBugs version 3.2.3 was the Gibbs sampling software used to compute the coefficients for the model

of TB. 70 000 iterations were run with three chains, giving 210 000 samples.

Table 2: Informative Priors

Coefficients:	Mean	Standard deviation
(Intercept)_alpha	-1.90	1.13
factor(hlsta)2_beta 1	1.52	0.17
factor(hiv)2_beta 2	-2.30	1.06
factor(con)2_beta 3	-1.11	0.25
factor(prov)2_beta 4	0.15	0.32
factor(prov)3_beta 5	0.10	0.33
factor(prov)4_beta 6	0.07	0.37
factor(prov)5_beta 7	0.27	0.28
factor(prov)6_beta 8	-0.20	0.38
factor(prov)7_beta 9	-0.44	0.38
factor(prov)8_beta 10	-0.27	0.40
factor(prov)9_beta 11	-0.77	0.45
factor(readen)2_beta 12	0.37	0.22
factor(gender)2_beta 13	-0.53	0.17
factor(edu)2_beta 14	-0.46	0.20
factor(edu)3_beta 15	-8.73	6.36
factor(hhdwlytp)2_beta 16	-0.23	0.23
factor(hhincome)2_beta 17	0.38	0.16
factor(marstt)2_beta 18	0.33	0.17
factor(marstt)3_beta 19	1.15	31.58
factor(alc)2_beta 20	-0.27	0.18

The posterior distributions of the model parameters generated from the sampled values, shown in Figure 1 reflected kernel densities.

The history of trace plots in Figure 2 shows convergence due to the overlapping of the three chains.

All BGR diagnostics plots were approaching a ratio of 1 indicating convergence for all variables. Figure 3 shows that there is no multicollinearity among the variables.

Comparison of Frequentist versus Bayesian Inference Approaches

The summary Statistics for FLR from R and for Bayesian non-informative and informative priors from OpenBugs version 3.2.3 are shown in Table 3.

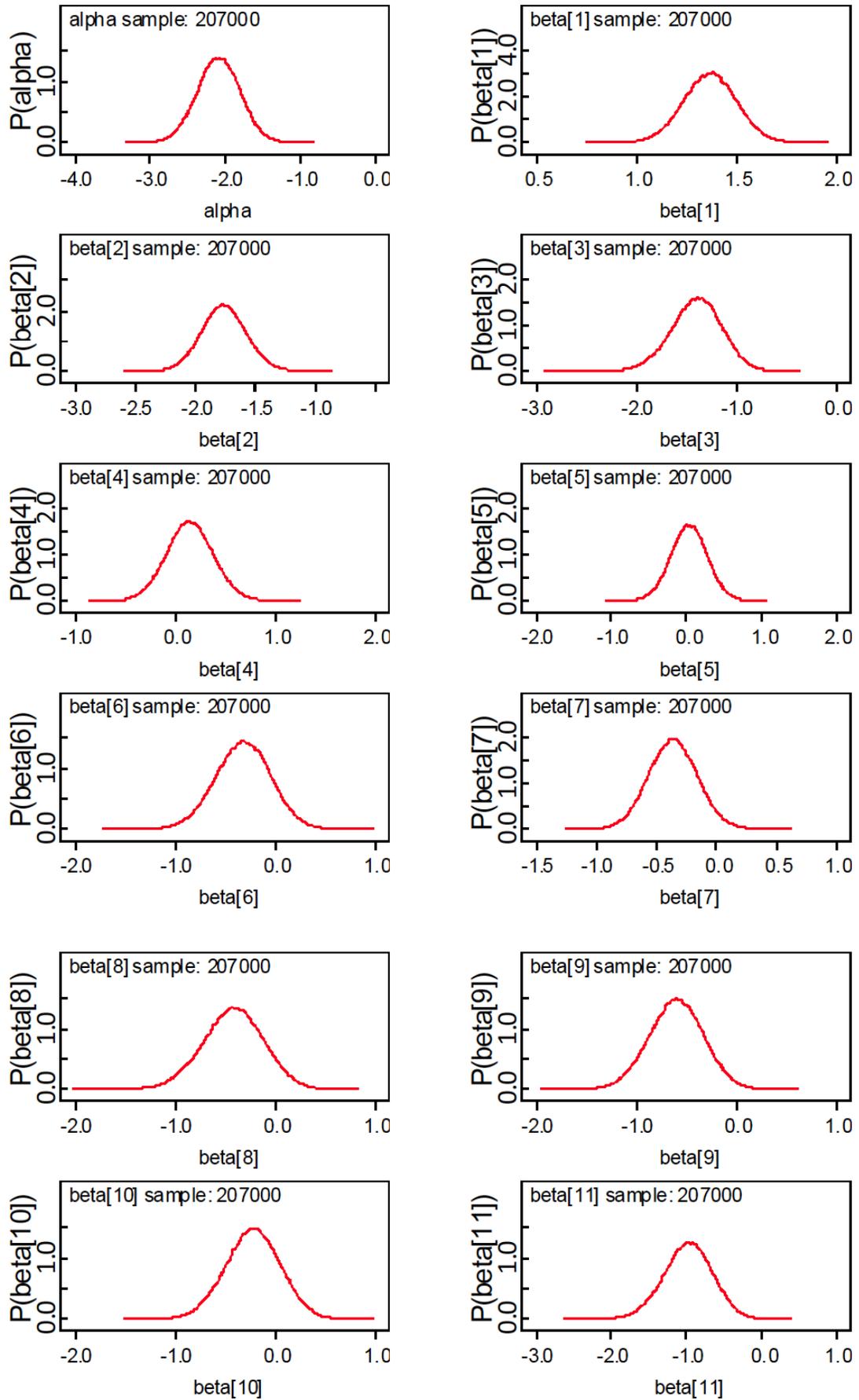
The parameter estimates obtained from the FLR and those obtained from the Bayesian non-informative prior are almost the same. The parameter estimates obtained from the Bayesian informative priors are slightly different in magnitude but the direction is the same which means the three models are explaining TB the same given the same variables. All the 95 % credible intervals for the significant variables in Table 4 do not include a zero and are therefore significant at 5% level of significance except marstt2. This is marital status, the divorced/widowed are significant with the informative priors. This means that the divorced/widowed are significantly different from the married. They have a higher chance of TB infection than the married.

CONCLUSION

From the descriptive statistics, we find that TB is more prevalent in males, the divorced/widowed, highest level of education Primary, Xhosa speaking, unemployed, age 45-59, do not exercise, value religion, drink alcohol, smoke, HIV positive and the illiterate. Households with no taped water, non-flushing toilets, uses other sources of energy that electricity, are below the poverty food line, have less than average income, are overcrowded, receive social grant, situated in informal settlements or are in traditional areas/farms have higher chances of having TB infected household members than their counterparts.

From the model that best describes the data, the individual risk factors are HIV status, English literacy, gender, education, marital status and alcohol. Though not risk factors, other factors associated with TB prevalence are self-reported health status, consulted about health, and exercise. There are some people who will be under TB treatment or still have the disease but perceiving their health status as good. This could be because the DOTS treatment strategy is working effectively to treat TB. The higher prevalence of TB in people who consulted about their health is possibly because most people consult at the clinic/doctor when they have fallen ill. Higher prevalence in people who do not exercise maybe be because once infected, TB patients will be physically unfit to exercise. Only two household variables are risk factors for TB, dwelling type and household income, thus households in informal settlements and households with below average income are more at risk of TB disease.

The results from the FLR are very similar to the Bayesian results with non-informative priors. The



(Figure 1). Continued.

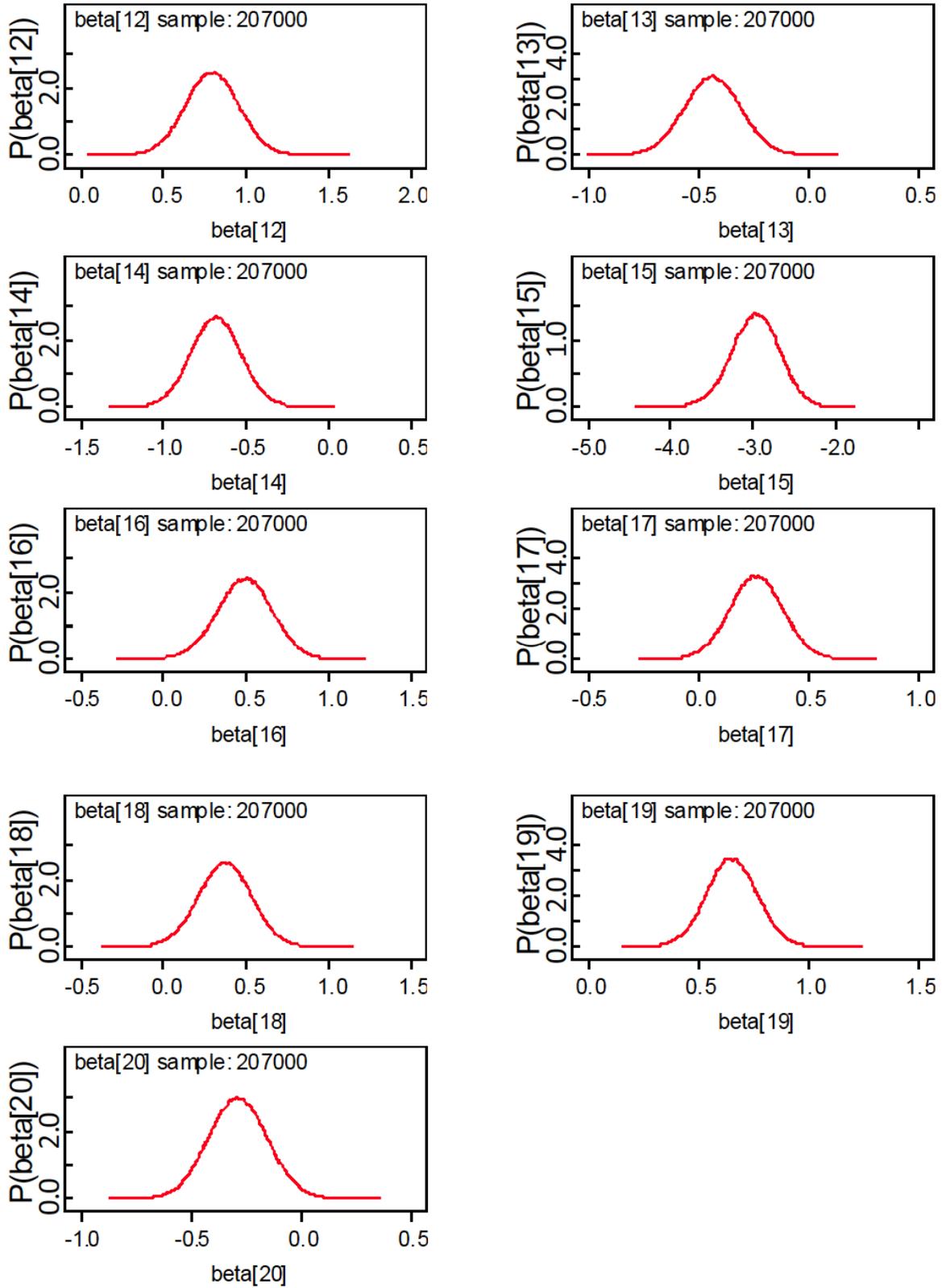
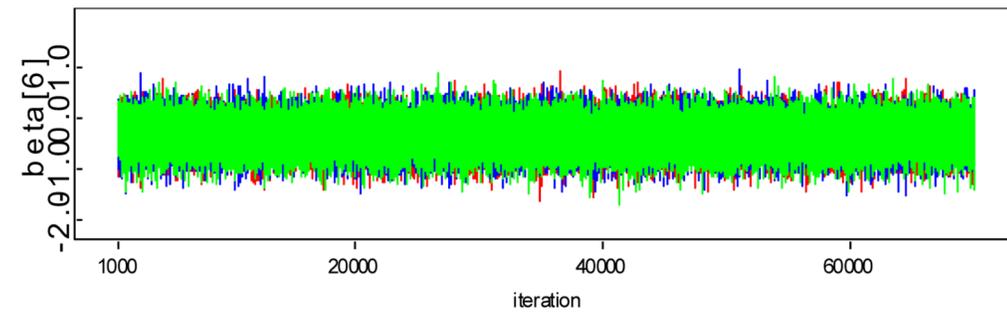
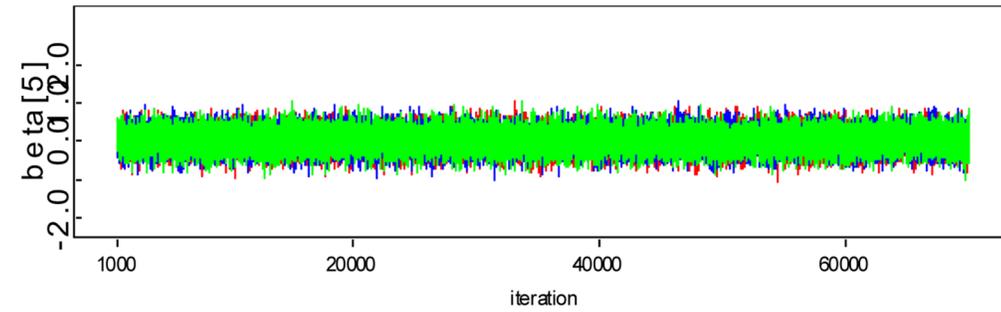
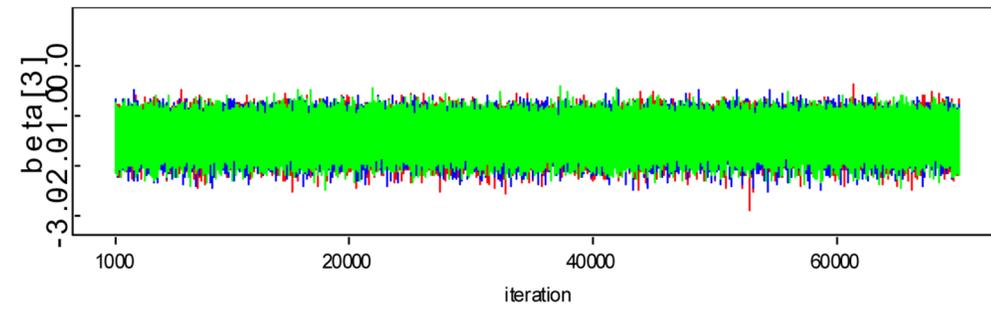
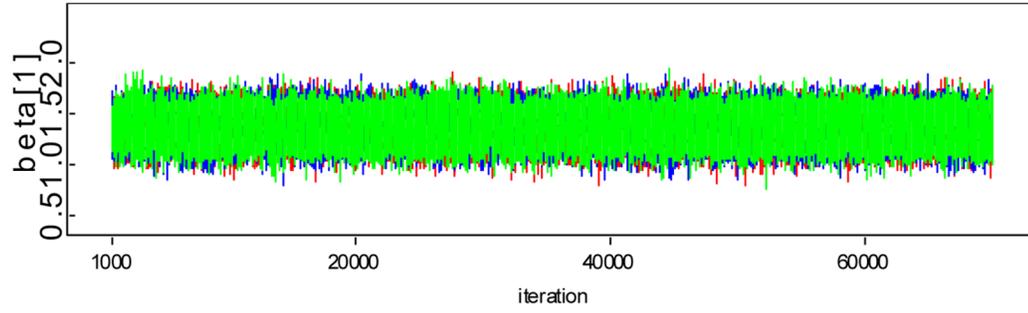
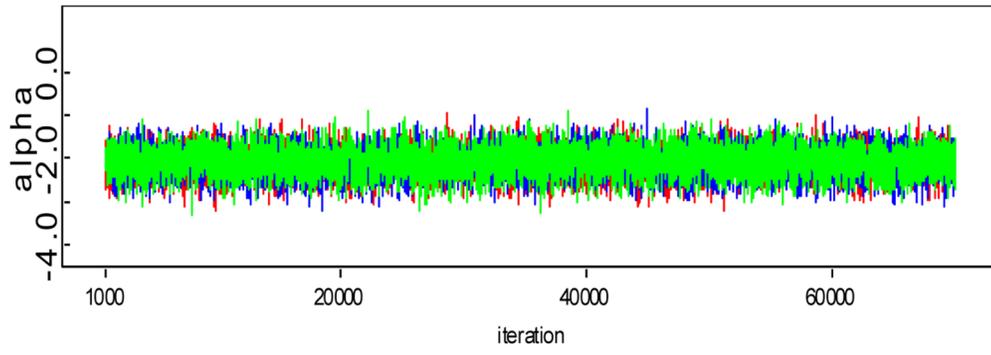
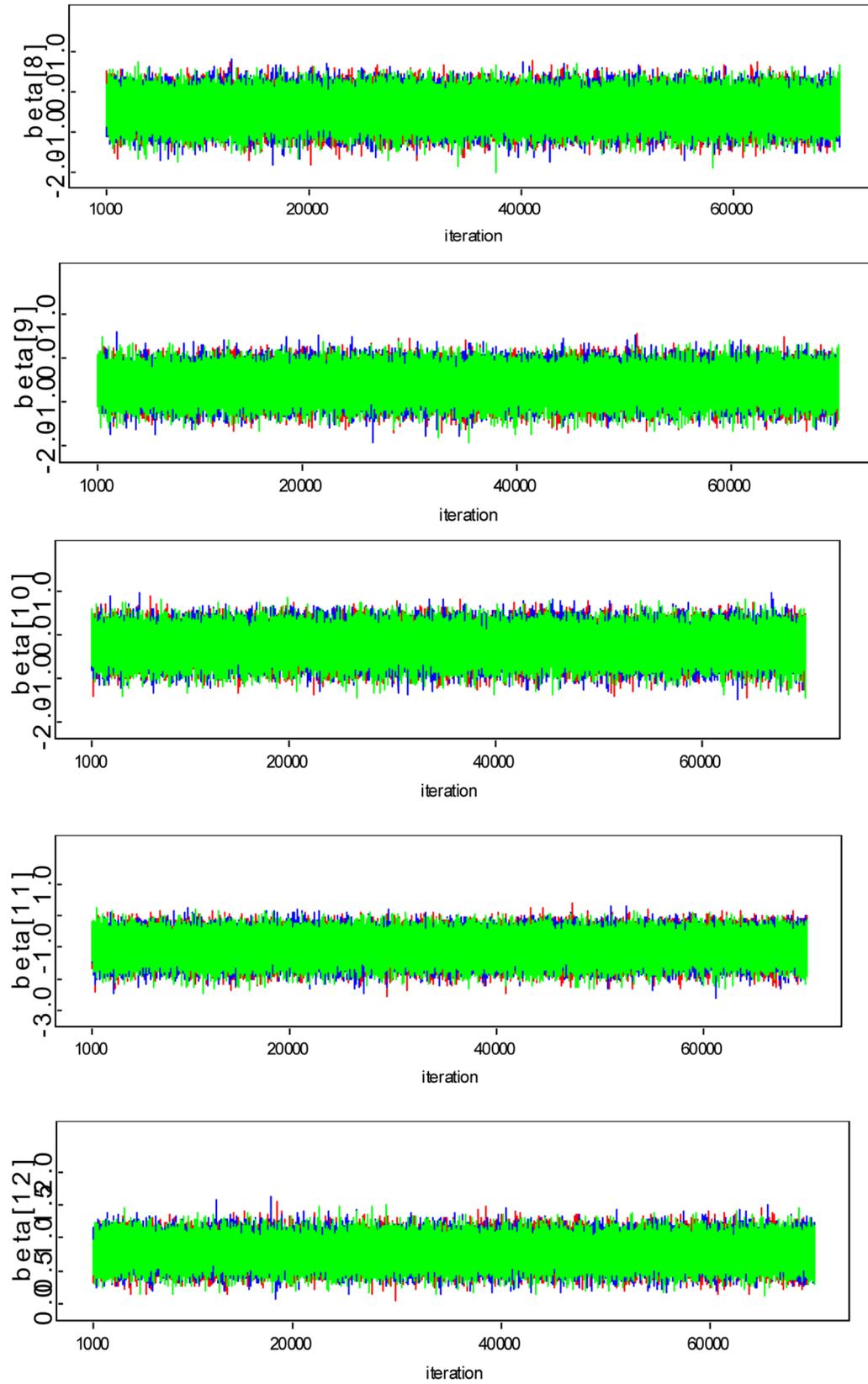


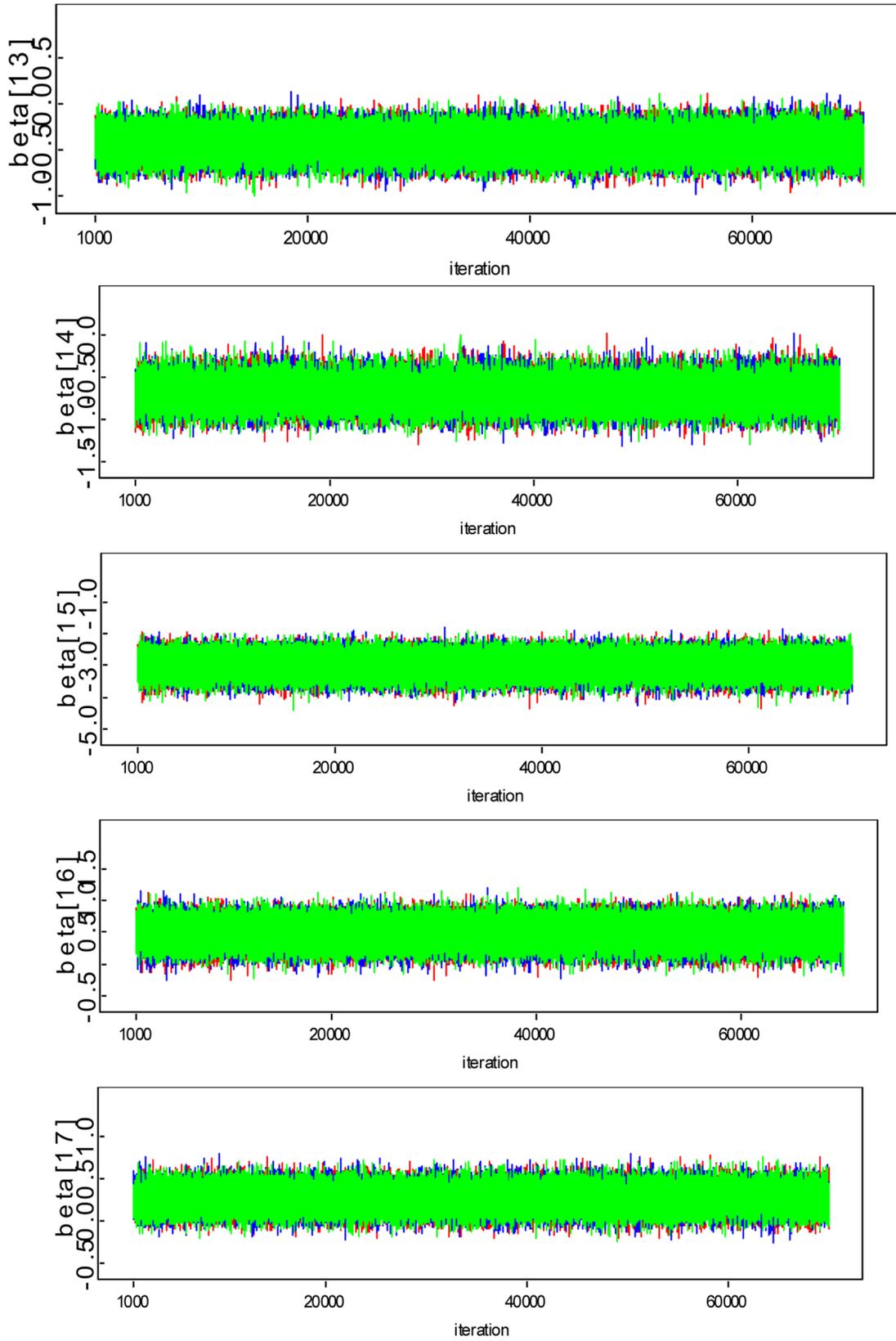
Figure 1: Posterior density distribution for informative priors.



(Figure 2). Continued.



(Figure 2). Continued.



(Figure 2). Continued.

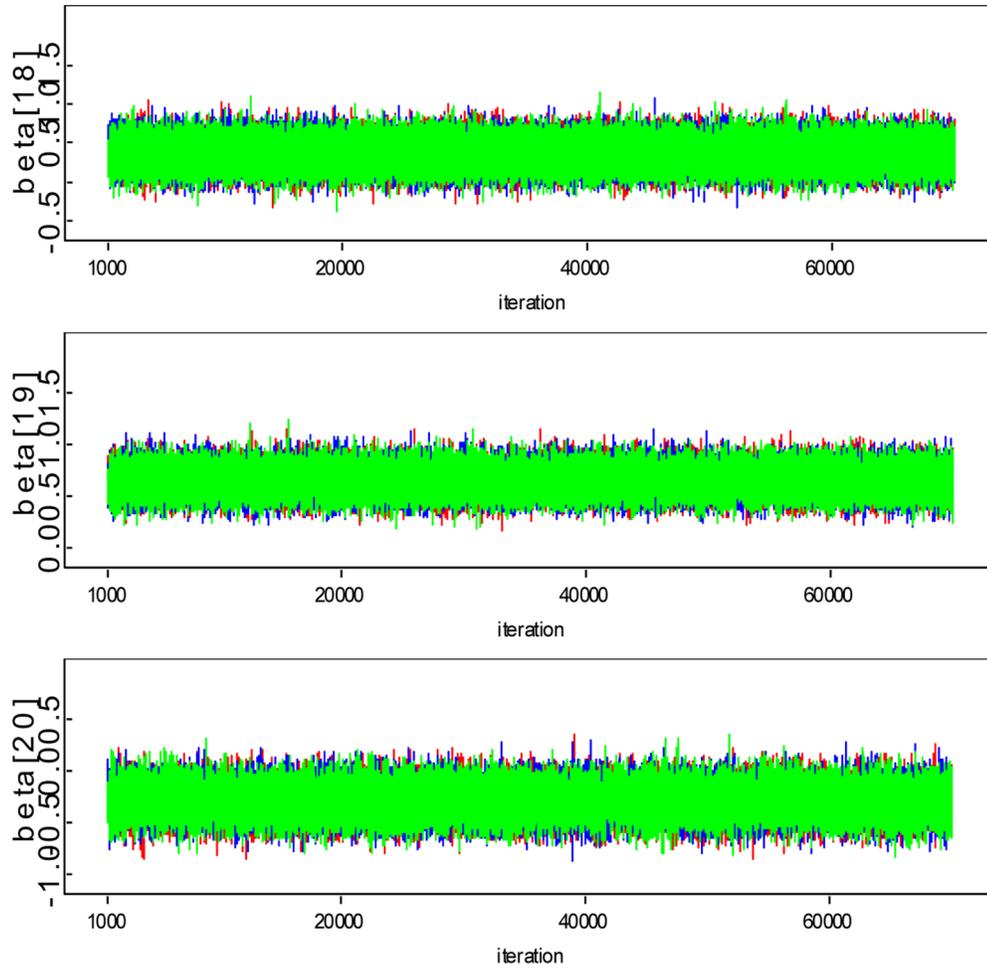


Figure 2: History of trace plots for the informative priors.

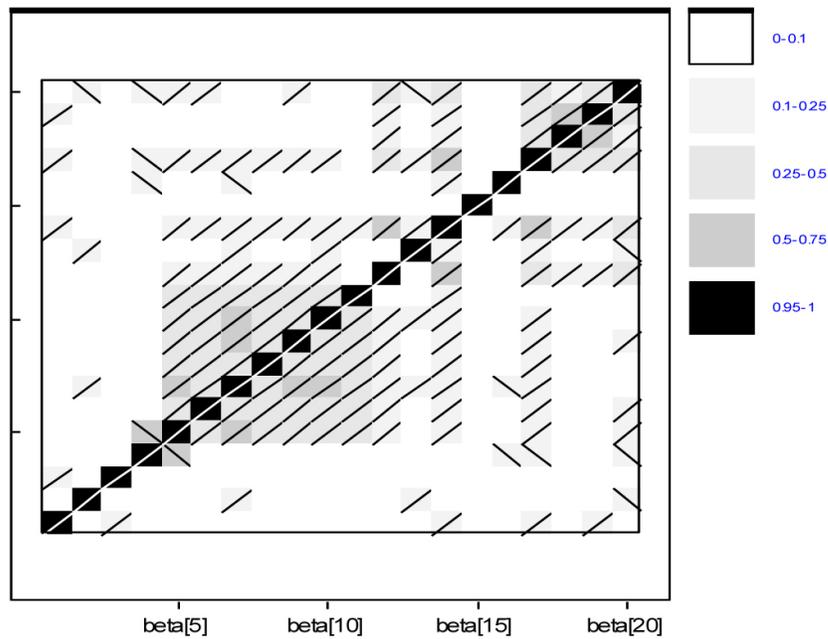


Figure 3: Cross-correlation matrix.

Table 3: Parameter Estimates and Standard Errors for FLR, Bayesian Non-Informative and Bayesian Informative Priors Respectively

	Frequentist		Bayesian Non-Informative Priors		Bayesian Informative Prior	
	Estimate	standard Error	mean	standard deviation	mean	standard deviation
(Intercept)	-1.94	0.30	-1.98	0.30	-2.09	0.29
beta [1]	1.24	0.13	1.24	0.13	1.37	0.13
beta [2]	-1.73	0.19	-1.72	0.19	-1.76	0.18
beta [3]	-1.35	0.25	-1.37	0.25	-1.40	0.25
beta [4]	0.15	0.21	0.13	0.23	0.15	0.23
beta [5]	0.04	0.24	0.03	0.25	0.03	0.24
beta [6]	-0.27	0.28	-0.28	0.28	-0.32	0.28
beta [7]	-0.42	0.21	-0.41	0.21	-0.36	0.20
beta [8]	-0.46	0.30	-0.48	0.31	-0.42	0.30
beta [9]	-0.62	0.27	-0.62	0.27	-0.60	0.27
beta [10]	-0.26	0.27	-0.27	0.28	-0.23	0.27
beta [11]	-1.01	0.33	-1.04	0.33	-0.98	0.32
beta [12]	0.39	0.18	0.39	0.18	0.79	0.16
beta [13]	-0.43	0.13	-0.43	0.13	-0.44	0.13
beta [14]	-0.57	0.15	-0.57	0.15	-0.68	0.15
beta [15]	-0.43	0.20	-0.43	0.21	-2.98	0.29
beta [16]	0.44	0.16	0.43	0.17	0.49	0.17
beta [17]	0.27	0.12	0.27	0.12	0.26	0.12
beta [18]	0.14	0.16	0.14	0.16	0.37	0.16
beta [19]	0.37	0.14	0.37	0.14	0.65	0.11
beta [20]	-0.27	0.13	-0.27	0.13	-0.29	0.13

Table 4: Bayesian Non-Informative Versus Informative Priors Credible Intervals

	Non –Informative priors					Informative priors				
	mean	standard deviation	MC	2.50%	97.50%	mean	standard deviation	MC	2.50%	97.50%
			error					error		
(Intercept)	-1.98	0.30	0.01	-2.56	-1.40	-2.09	0.29	0.00	-2.65	-1.54
beta [1]	1.24	0.13	0.00	0.98	1.5	1.37	0.13	0.00	1.11	1.63
beta [2]	-1.72	0.19	0.00	-2.08	-1.35	-1.76	0.18	0.00	-2.12	-1.40
beta [3]	-1.37	0.25	0.00	-1.9	-0.89	-1.40	0.25	0.00	-1.92	-0.93
beta [4]	0.13	0.23	0.00	-0.33	0.6	0.15	0.23	0.00	-0.31	0.61
beta [5]	0.03	0.25	0.00	-0.46	0.52	0.03	0.24	0.00	-0.44	0.51
beta [6]	-0.28	0.28	0.00	-0.84	0.26	-0.32	0.28	0.00	-0.88	0.21
beta [7]	-0.41	0.21	0.00	-0.82	0.01	-0.36	0.20	0.00	-0.76	0.04
beta [8]	-0.48	0.31	0.00	-1.1	0.1	-0.42	0.30	0.00	-1.03	0.15

(Table 4). Continued.

	Non –Informative priors					Informative priors				
	mean	standard deviation	MC error	2.50%	97.50%	mean	standard deviation	MC error	2.50%	97.50%
beta [9]	-0.62	0.27	0.00	-1.17	-0.1	-0.60	0.27	0.00	-1.14	-0.08
beta [10]	-0.27	0.28	0.00	-0.82	0.27	-0.23	0.27	0.00	-0.77	0.30
beta [11]	-1.04	0.33	0.00	-1.7	-0.4	-0.98	0.32	0.00	-1.63	-0.37
beta [12]	0.39	0.18	0.00	0.04	0.73	0.79	0.16	0.00	0.48	1.11
beta [13]	-0.43	0.13	0.00	-0.68	-0.18	-0.44	0.13	0.00	-0.69	-0.18
beta [14]	-0.57	0.15	0.00	-0.86	-0.26	-0.68	0.15	0.00	-0.97	-0.39
beta [15]	-0.43	0.21	0.00	-0.84	-0.03	-2.98	0.29	0.00	-3.56	-2.43
beta [16]	0.43	0.17	0.00	0.1	0.75	0.49	0.17	0.00	0.16	0.81
beta [17]	0.27	0.12	0.00	0.04	0.5	0.26	0.12	0.00	0.03	0.50
beta [18]	0.14	0.16	0.00	-0.19	0.45	0.37	0.16	0.00	0.05	0.68
beta [19]	0.37	0.14	0.00	0.1	0.66	0.65	0.11	0.00	0.43	0.88
beta [20]	-0.27	0.13	0.00	-0.53	-0.02	-0.29	0.13	0.00	-0.55	-0.02

Bayesian model with informative priors gives slightly different results from the two models.

RECOMMENDATIONS

We recommend that in addition to the existing TB interventions, the government must also concentrate on improving living conditions in informal settlements by providing access to clean taped water. Low socio-economic communities need continued support to improve their economic status. Continued efforts to reduce illiteracy rates and increase the proportion of people with Secondary/ Tertiary will also help eradicate TB. More resources must be channelled to Eastern Cape as it is the one most hit by TB.

REFERENCES

[1] WHO report (2011); available from: http://www.who.int/whosis/whostat/EN_WHS2011_Full.pdf?ua=1

[2] TB Statistics for South Africa – National & provincial available from <http://www.tbfacts.org/tb-statistics-south-africa/#sthash.7MG2pSTA.dpuf>

[3] WHO End TB Strategy 2015; available from http://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1

[4] TB Statistics for South Africa-national and provincial www.tbfacts.org/tb-statistics-south-africa

[5] Global TB strategy 2015; available at: http://pdf.usaid.gov/pdf_docs/PNADR735.pdf

[6] Cramm JM, Koolman X, Møller V, Nieboer AP. Socioeconomic status and self-reported tuberculosis: a multilevel analysis in a low-income township in the Eastern

Cape, South Africa. Journal of Public Health in Africa Research and Practise 2011; 2(2): 143-146. <https://doi.org/10.4081/jphia.2011.e34>

[7] Harling G, Ehrlich R, Myer L. The Social Epidemiology of Tuberculosis in South Africa: A Multilevel Analysis. Social Science & Medicine 2008; 66(2): 492-505. <https://doi.org/10.1016/j.socscimed.2007.08.026>

[8] Janssens J, Rieder H, An ecological analysis of incidence of tuberculosis and per capita gross domestic product. Eur Respir J 2008; 32(5): 1415-1416. <https://doi.org/10.1183/09031936.00078708>

[9] Mozolo T, Mwambi H, Zuma K. Bayesian approach in estimating risk determinants of infectious diseases. Presented at the 51st Annual South African Statistical Association (SASA) Conference, University of Pretoria, Pretoria, South Africa 27-31 October 2008.

[10] Lesaffre E, Lawson B. Bayesian Biostatistics, John Wiley and Sons 2012.

[11] Bolstad WM. Introduction to Bayesian Statistics, New jersey, USA, John Wiley and Sons 2007.

[12] Mohammed OMM. Statistical methods for analysing complex survey data: An application to HIV/AIDS in Ethiopia. PhD, 2013.

[13] Leibbrandt M, Woolard I, De Villiers L. Methodology: Report on NIDS Wave 1. Cape Town: University of Cape Town, 2009. <http://www.nids.uct.ac.za/publications/technical-papers/108-nids-technical-paper-no1/file>

[14] Akaike H. Information theory and an extension of the maximum likelihood principle. 1973. Proc. 2nd Inter. Symposium on Information Theory 267-281, Budapest.

[15] Siddiqui AA, Siddiqui JS, Wasay M, Azam SI, Ahmed A. A dynamical study of risk factors in intracerebral haemorrhage using the multivariate approach. International Journal of Statistics in Medical Research 2013; 2: 23-33.

[16] Khaial FB, Bodalal Z, Elramli A, Elkhwsy F, Eltaguri A, Bendardaf R. A study of risk factors for breast cancer in a primary oncology clinic in Benghazi-Libya. Internation Journal of Statistics in Medical Research 2015; 4(1).

- [17] Ryan TP. Modern regression methods. New York, JohnWiley and Sons 2009.
- [18] Holmes CC, Caron F, Griffin JE, Stephens DA. Two-sample Bayesian Nonparametric Hypothesis Testing *Bayesian Anal* 2015; 10(2): 297-320.
<https://doi.org/10.1214/14-BA914>
- [19] Congdon P. Bayesian methods and Bayesian estimation, in *Applied Bayesian Modelling*, John Wiley & Sons, Ltd, Chichester, UK 2014.
<https://doi.org/10.1002/9781118895047.ch1>
- [20] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990; 85(40): 398-409.
<https://doi.org/10.1080/01621459.1990.10476213>

Received on 22-11-2016

Accepted on 12-01-2017

Published on 28-02-2017

<https://doi.org/10.6000/1929-6029.2017.06.01.4>

© 2017 Dhlakama and Lougue; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.