Scientific
Research
Publishing

Spatial Modeling and Mapping of Tuberculosis Using Bayesian Hierarchical Approaches

Abdul-Karim Iddrisu[1], Yaw Ampem Amoako[2]

[1]Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa
[2]Department of Medicine, Komfo Anokye Teaching Hospital, Kumasi, Ghana
Email: karim@aims.ac.za, yamoako2002@yahoo.co.uk

Abstract

Global spread of infectious disease threatens the well-being of human, domestic, and wildlife health. A proper understanding of global distribution of these diseases is an important part of disease management and policy making. However, data are subject to complexities by heterogeneity across host classes. The use of frequentist methods in biostatistics and epidemiology is common and is therefore extensively utilized in answering varied research questions. In this paper, we applied the hierarchical Bayesian approach to study the spatial distribution of tuberculosis in Kenya. The focus was to identify best fitting model for modeling TB relative risk in Kenya. The Markov Chain Monte Carlo (MCMC) method via WinBUGS and R packages was used for simulations. The Deviance Information Criterion (DIC) proposed by [1] was used for models comparison and selection. Among the models considered, unstructured heterogeneity model perfumes better in terms of modeling and mapping TB RR in Kenya. Variation in TB risk is observed among Kenya counties and clustering among counties with high TB Relative Risk (RR). HIV prevalence is identified as the dominant determinant of TB. We find clustering and heterogeneity of risk among high rate counties. Although the approaches are less than ideal, we hope that our formulations provide a useful stepping stone in the development of spatial methodology for the statistical analysis of risk from TB in Kenya.

Keywords

1. Introduction

The global spread of infectious diseases threatens the well-being of human population as well as that of domestic and wild animals. Therefore, a proper understanding of the pathways and global spread of communicable and infectious diseases is an important aspect of disease management and policy making. However, the data collected with the object of understanding these patterns are subjected to complications brought about by heterogeneity that can be of spatial or non-spatial in nature [2] [3]. Ignoring these complexities is likely to mislead inference which may results in erroneous conclusion [2] [3]. Disease data can be case-event or counts from non-over-lapping regions. Since the data provided for this study only have county level spatial resolution, the approach will be to analyses the data as regional or county specific count data.

There exists a vast amount of literature concerning the development and application of disease mapping approaches [4]-[8]. Spatial distribution of a disease is often understood through application of statistical methods to the data and creating maps that visually describes spatial variation of disease risk [3]. However, disease counts maps are subjected to numerous problems. One such problem is the Modifiable Areal Unit Problem (MAUP), which occurs when inference at the areal level differs from that which is observed at the basic observational unit. This is likely to change conclusions drawn from a study of a count data. The MAUP can be addressed by scaling up to ensure smoothing or averaging of data and making inference at high aggregate level than that used in the analysis. MAUP can also be addressed by scaling down to enable inference at lower level than that used in the analysis. Multi-scale Analysis can also be used to addressed MAUP. This analysis concerns spatial units that are completely matched when aggregated [9]. Also differences in population between regions result in differences in variance of regional estimates. This problem is addressed by employing a hierarchical Bayesian model that smooths the risk from neighboring regions and clearly accounts for population difference by using a Poisson distribution for outcomes.

Bayesian methods are widely used in disease mapping. [10] applied the the Empirical Bayes (EB) methods for smoothing a map of lip cancer rates. They assumed a multivariate normal for the log relative risks and allowed for spatial correlation via conditional autoregressive model. Their model could not be considered to be a "fully Bayesian", since a quadratic approximation was used for the likelihood and this did not account for the uncertainty in the estimates of the hyper-parameters. [11] was the first example of fully Bayesian disease mapping. They used the convolution prior model described in Section 6.1 to model the log relative risk. They found that the model shrunk extreme disease rates towards the mean and detected spatial association that was apparent in the raw data. According to [11], the fully Bayesian model produced more accurate estimates than that produced by Clayton and Kaldor approach.

[12] ecological study applied Bayesian hierarchical regression model to evaluate the urban spatial and spatio-temporal distribution of TB in Rilirão Preto, state of São Paulo, Southeast Brazil between 2006-2009 and to evaluate TB risk determinants. The study reveals that TB rates are correlated with measures of income, education and social vulnerability. They state that complex relationship may exist between TB incidence and a wide range of environmental and intrinsic factors, which need to be studied in future research. [13] applied both the Bayesian approach and the generalized mixed model to produce smooth relative risk maps of TB and to model relationship between TB new cases and national TB control program indicators. Their study discovered that high TB risk areas were clustered and TB distribution found to be associated with the number of patients lost to follow-up and the number house holds with more than one case. [14] study on assessing the prevalence of TB in New York from 1970-2000 using Bayesian analysis approach stated that decline in TB incidence could probably be as a result of good control programs and raised in TB prevalence could be attributed to social disruptions such as homelessness, drug abuse, poverty, and overcrowding. Their study confirmed that increase in TB is mainly due to HIV epidemic. [15] study of TB pattern in India, using the Bayesian conditional autoregressive model revealed that north-eastern states have high risk of TB than other regions.

In this paper our focus is to propose best fitting Hierarchical Bayesian approaches for modeling and mapping relative risk tuberculosis in Kenya. Specifically, we determine suitable spatial and non-spatial models for modeling TB in Kenya.

In Section 2, we describe the data used followed by a Bayesian modeling framework in Section 3. In Section 4, we describe two relative risk detection methods. In Section 5, we described Bayesian non-spatial models and spatial models in Section 6 for disease mapping. We give discussion and conclusion in Section 7.

2. Data Description

The data used in this study is routine data from Kenya Demographic and Health Surveys (DHS). The data contain records of Kenya's population size, tuberculosis cases, and some suspected determinants of tuberculosis for each period from 2002-2009 and for each 67 districts. To study the risk of TB infection in each county, the data from the 67 districts were aggregated to provide county level summaries. Some of the determinants of TB that were recorded are: HIV prevalence, poverty prevalence, illiteracy, population less then 5 km to health facility, firewood, altitude and mean house hold size. Summary Statistics for these determinants of TB are shown in **Table 1**.

3. Bayesian Modeling Framework

Bayesian methods define the posterior distribution which is the distribution of the parameter(s) given the observed data. This implies that we require a likelihood function for the observed data and a prior distribution for the unknown parameters. Let $y_i, i = 1, \cdots, n$ be a random variable with probability density function $P(y_i | \boldsymbol{\vartheta})$, where $\boldsymbol{\vartheta} = (\vartheta_1, \cdots, \vartheta_p)$ is a vector of relative risk parameters. The likelihood function of y_i is defined as

$$P(\boldsymbol{y} | \boldsymbol{\vartheta}) = \prod_{i=1}^{n} P(y_i | \boldsymbol{\vartheta}). \tag{1}$$

Equation (1) is based on the assumption that the sample values of $\boldsymbol{y} = (y_1, \cdots, y_n)'$ given the parameters $\boldsymbol{\vartheta}$ are independent [9]. Once the data model (likelihood function) is chosen, a Bayesian analysis requires the assertion of a prior distribution for the unknown model parameters. The prior distribution can be viewed as representing the current state of knowledge, or current description of uncertainty, about the model parameters prior to data being observed. A prior distribution is a distribution assigned to the parameter $\boldsymbol{\vartheta}$ before the data y_i are observed [9]. All parameters within the Bayesian models are stochastic and assigned appropriate prior distribution [9]. Given a single parameter, ϑ, the the prior distribution is denoted as $P(\vartheta)$, while for a parameter vector, $\boldsymbol{\vartheta}$, the joint prior distribution is denoted as $P(\boldsymbol{\vartheta})$.

The posterior distribution is a probability distribution of the parameters given the data. The posterior distribution which is proportional to the product of the likelihood function and the prior distribution is defined as

$$P(\boldsymbol{\vartheta} | \boldsymbol{y}) = \frac{P(\boldsymbol{y} | \boldsymbol{\vartheta}) P(\boldsymbol{\vartheta})}{\int_p L(\boldsymbol{y} | \boldsymbol{\vartheta}) P(\boldsymbol{\vartheta}) d_{\boldsymbol{\vartheta}}}, \tag{2}$$

where $\int_p L(\boldsymbol{y} | \boldsymbol{\vartheta}) P(\boldsymbol{\vartheta}) d_{\boldsymbol{\vartheta}}$ is called the normalizing constant. It have been shown that posterior distribution of the parameters (2) can be written as

$$P(\boldsymbol{\vartheta} | \boldsymbol{y}) \propto P(\boldsymbol{y} | \boldsymbol{\vartheta}) P(\boldsymbol{\vartheta}), \tag{3}$$

Table 1. Summary statistics for TB determinants.

Variables	No. of counties	Mean	SD	Median	Min	Max	95% CI
HIV prevalence (%)	47	4.289	2.797	3.800	1.000	16.430	(2.950, 4.700)
Proportion of poor	47	0.5196	0.184	0.5013	0.1157	0.9434	(0.3778, 0.6369)
Illiteracy (%)	47	24.47	19.958	16.00	2.80	77.30	(12.10, 29.80)
Household 5 km away from hospital (%)	47	77.76	16.399	80.80	19.20	99.00	(72.05, 86.72)
Firewood (%)	47	78.52	20.175	84.60	1.80	96.70	(74.95, 90.65)
Altitude (m)	47	1361	602.2214	1432	151	2274	(1138, 1813)
Mean house hold size	47	5.383	0.799	5.250	3.800	6.900	(4.775, 6.050)

from which parameter estimates are drawn using Markov Chain Monte Carlo (MCMC) vis Gibbs sampling [16]. True parameter estimates are obtained when the the Markov chain converges. Convergence diagnostics for each parameter estimate under each implementation is shown.

4. Cluster Detection

This section discusses methods used in disease mapping to detect elevated risk. One such methods is the use of the posterior $P(\vartheta \mid y)$ measures in posterior distribution (3). Another approach for risk detection is the use of the exceedence probability which relies on the posterior $P(\vartheta \mid y)$ measures for cluster detection [9]. The most commonly criteria for cluster detection method is the exceedence probability in relation with the relative estimates for individual areas or counties [17]. The exceedence probability is defined as the probability that the relative risk ϑ exceeds some threshold level $\Bbbk$, defined by $P(\vartheta > \Bbbk)$. Exceedence probability is computed from the posterior sample values $\{\vartheta_i^g\}\{g = 1, 2, \cdots, G\}$ through $\hat{P}(\vartheta_i > \Bbbk) = \frac{1}{G}\sum_{g=1}^{G} I(\vartheta_i^g > \Bbbk)$, where

$$I(\vartheta_i^g > \Bbbk) = \begin{cases} 1 & \text{if} \quad \vartheta_i^g > \Bbbk \\ 0 & \text{Otherwise.} \end{cases} \tag{4}$$

In evaluating $P(\vartheta_i > \Bbbk)$, $\Bbbk$ and the threshold probability must be chosen such that $P(\vartheta_i > \Bbbk) > k$. By convention, k can take the values of 0.95, 0.975, 0.99, etc. [18]. According to [9], exceedence probability is only capable of detecting hot spot cluster and does not consider any other information concerning possible forms of cluster or even neighborhood information. [9] defined Hot spot as any area displaying "excess" or "unusual" risk. According [9], Hossain and Lawson have made some attempts to enhance the exceedence probability by inclusion of neighborhoods. They stated that, for the neighborhood of the ith area defined as a_i and the number of neighbors as n_i, then

$$\bar{A}_i = \frac{\sum_{j^*=0}^{n_i} A_{ij^*}}{(n_i + 1)} \quad \text{where } A_{ij^*} = P(\vartheta_{j^*} > \Bbbk) \,\forall j^* \in a_i \text{ and } A_{i0} = P(\vartheta_i > \Bbbk) \tag{5}$$

These measures can be used to detect different forms of clustering [9]. However, the usefulness of the execeedence probability depends on the model that has been fitted to the data and that any poorly fitting model will not demonstrate exceedence relate to clustering [9].

5. Bayesian Hierarchical Non-Spatial Models for Disease Mapping

This section presents two non-spatial models used in disease modeling and mapping. These are the Poisson-Gamma (PG) and the Poisson Log-Normal (LN) models. These models are often used to model small area count data and are appropriate when there is relatively low count of disease and the target population is relatively large in each small area. We present the PG model first followed by the LN model.

5.1. The Poisson-Gamma Model (PG)

The Poisson-Gamma model for relative risk estimation uses a gamma prior distribution for the relative risk combines with the Poisson likelihood function for the disease counts which gives a gamma posterior distribution for the relative risk. The Poisson-Gamma model is widely used in disease mapping to account for extra variability in the data through the prior distribution [18].

5.2. Model Description

Suppose that the unknown risk of TB in region i is given as $\vartheta_i, i = 1, 2, \cdots, n$. Let y_i and N_i denote the number of TB cases and the population at risk respectively in region i. The expected number of TB cases in region i can then be written as $E_i = rN_i$, where $r = \dfrac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} N_i}$ is the overall disease risk in the study population.

Under the frequentist paradigm, we assume that $y_i \sim Poisson(E_i \vartheta_i)$. Based on the sample $\boldsymbol{y} = \{y_1, \cdots, y_n\}$, the likelihood function and the corresponding log-likelihood function are expressed as

$$\ell(\vartheta_i) = \prod_{i=1}^{n} \frac{\exp(-E_i \vartheta_i)(E_i \vartheta_i)^{y_i}}{y_i!} = P(\boldsymbol{y}, \boldsymbol{E} \mid \boldsymbol{\vartheta}) \tag{6}$$

and

$$\ln \ell(\vartheta_i) = -\sum_{i=1}^{n} E_i \vartheta_i + \sum_{i=1}^{n} y_i \ln(E_i \vartheta_i) - \prod_{i=1}^{n} y_i! \tag{7}$$

The maximum likelihood estimator $\hat{\vartheta}_i$ of ϑ_i is obtained via $\dfrac{\partial(\ln \ell(\vartheta_i))}{\partial \vartheta_i} = 0$ and is given by

$$\hat{\vartheta}_i = \frac{y_i}{E_i}. \tag{8}$$

This estimator, $\hat{\vartheta}_i$ is referred to as the standardized mortality ratio in region i. Under the Bayesian paradigm, assume that $y_i \sim Poisson(E_i \vartheta)$, where μ is $\mu_i = E_i \vartheta$ is the Poisson mean and $\vartheta \sim Gamma(a,b)$ has a gamma distribution with shape and scale parameters a and b respectively. The likelihood function for y_i is denoted by

$$\ell(\vartheta) = \prod_{i=1}^{N} \frac{(E_i \vartheta)^{y_i} \exp(-E_i \vartheta)}{y_i!} = P(\boldsymbol{y}, \boldsymbol{E} \mid \vartheta). \tag{9}$$

and prior distribution for ϑ is denoted by

$$P(\vartheta \mid a,b) = \frac{(\vartheta)^{a-1}}{\Gamma(a)b^a} \exp(-\vartheta b), \vartheta, a, b > 0. \tag{10}$$

5.3. Parameter Estimation

We used Bayesian hierarchical methods for parameters estimation in the Poisson-Gamma model. That is, if in addition, a and b are given prior distributions such that $a \mid \omega \sim P(a \mid \omega)$ and $b \mid \boldsymbol{\phi} \sim P(b \mid \boldsymbol{\phi})$, where $P(a \mid \omega)$ and $P(b \mid \boldsymbol{\phi})$ are the hyperprior distribution with hyper-parameters ω and $(\phi_a, \phi_b) \in \boldsymbol{\phi}$ for a and b respectively, then we can obtain parameters using Bayesian hierarchical methods. This is a second stage hierarchical modeling using the Poisson-Gamma model.

In this paper, we defined $P(a \mid \omega) = \omega \exp(-\omega a)$ (as exponential distribution) and $P(b \mid \phi_a, \phi_b)$ as a gamma distribution. Therefore, the posterior distribution is given by

$$P(\vartheta, a, b, \mid \boldsymbol{y}, \boldsymbol{E}) \propto P(\boldsymbol{y}, \boldsymbol{E} \mid \vartheta, a, b,) P(\vartheta) P(a \mid \omega) P(b \mid \boldsymbol{\phi}). \tag{11}$$

Parameters estimation of the Poisson-Gamma model was carried out using MCMC via Gibbs Sampling. Convergence of the Chain occurs at 40,000 iterations after a burnin period of 1000 sample and thinning of every 30$^{\text{th}}$ element of the sample. **Figure 1** presents the MCMC convergence diagnostics plots.

5.4. Markov Chain Monte Carlo Diagnosis

Figure 1 presents Gelman and Rubin convergence diagnostics of the Poisson-Gamma model: Column-wise from the top left, **Figures 1(a)-(j)** are trace plots for a, b, the mean and variance respectively. **Figures 1(b)-(k)** are posterior marginal density plots for b, a, the mean and the variance respectively. **Figures 1(c)-(l)** are autocorrelation plots for b, a, the mean and the variance respectively. The Gelman and Rubin trace plots show the convergence of the two parallel chains (Chains with different initial values). "Vanishing" autocorrelation function (ACF) plots show that there is low correlation among parameters that constitute the chain. More satisfactory kernel density plots for parameters of interest would more bell-shaped or symmetric. The density plots for the parameters show that convergence of the chain has reached.

The posterior statistics of the Poisson-Gamma model are shown in **Table 2**. **Table 2** shows that the mean of
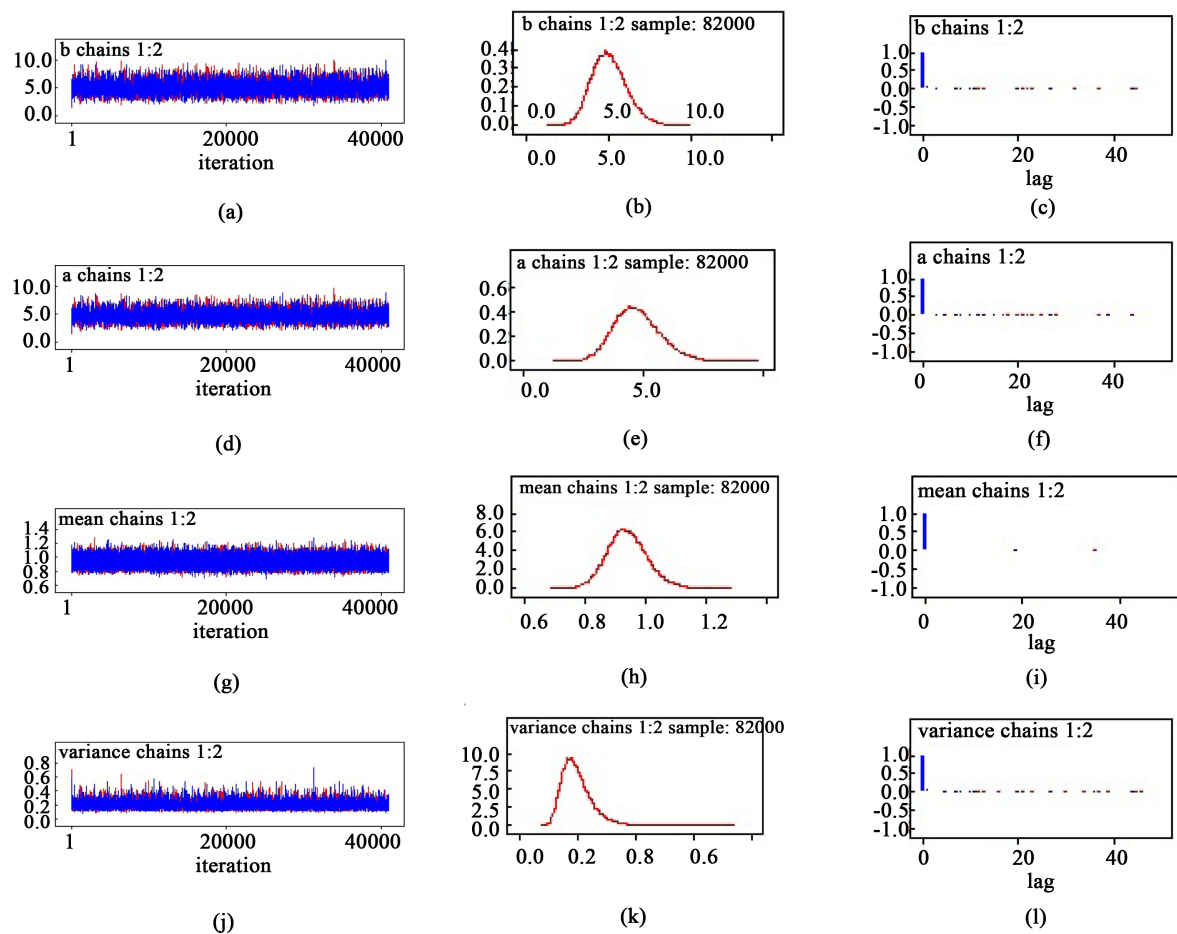
Figure 1. Poisson-gamma model: convergence diagnosis of Markov Chain Monte Carlo.

Table 2. Posterior statistics of the poisson-gamma model.

Parameters	Posterior Means	Credible Region
a	4.717	(3.089, 6.707)
b	5.046	(3.21, 7.29)
mean	0.933	(0.8192, 1.075)
variance	0.1955	(0.122, 0.3142)
$\bar{D}$	575,755	-
pD	46.946	-
DIC	622.701	-

the posterior relative risk is 0.93 (95% credible interval = 0.82 - 1.08). The posterior mean is approximately the same as the mean of the standardized mortality ratio 0.93 (95% credible interval = 0.62 - 1.05) (Equation (8)). The standard deviation of the relative risk, 0.42 (95% credible interval = 0.35 - 0.56), is lower than the standardized mortality ratio's standard deviation 0.49 (95% credible interval = 0.62 - 1.05). Thus their standard deviation has been reduced by 82% by the Poisson-Gamma model. The significance of the variance indicates variation in risk among counties. In a situation of rare cases, standard deviation of the Poisson-Gamma model is expected to be much lower than that of the standardized mortality ratio [18].

Figure 2 shows standardized mortality ratio for TB prevalence in the counties of Kenya for 2002-2009. The
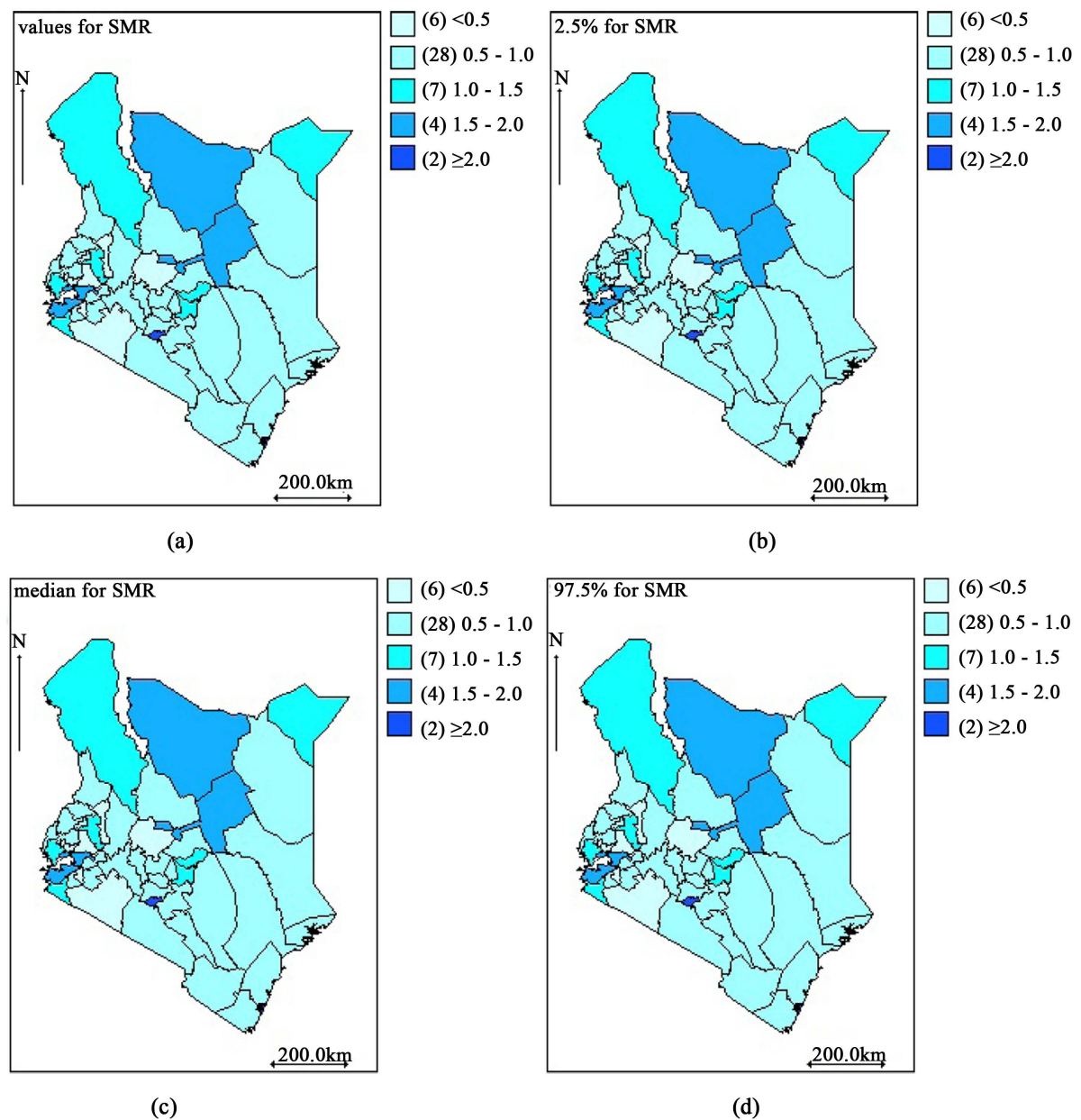
Figure 2. Kenya county level standardized mortality ratio's maps: (a) The mean of the SMR and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

SMRs vary around their mean, 0.93 with standard deviation, 0.49 (as discussed in **Table 2**). There is some suggestion of high TB prevalence in the North, West, North-West and Central counties of Kenya and low TB prevalence in the South-East counties except Mombasa (SMR > 2.0).

From **Figure 3**, we observed high risk of TB prevalence in the North, West, North-West and Central counties of Kenya and low risk in the South-East counties except Mombasa. Nairobi and Mombasa have the highest relative risk (RR > 2.0) and Laikipia, Nandi, Narok, Nyamira, and Vihiga have the lowest risk (RR < 0.5). The mean of the posterior relative risk and the SMR are the same. The range of the posterior relative risk of the Poisson-Gamma remains the same as the SMR, each having lowest relative risk estimated at 0.40 and the highest risk at 2.38. Variability in risk remains the same due to abundant of information or data.

The map of the exceedence probability in **Figure 4** revealed 13 counties that exhibit high risk of TB above the national risk (RR > 1). These counties are: Nairobi, Mombasa, Kisumu, Turkan, Migori, Homa bay, Uasin
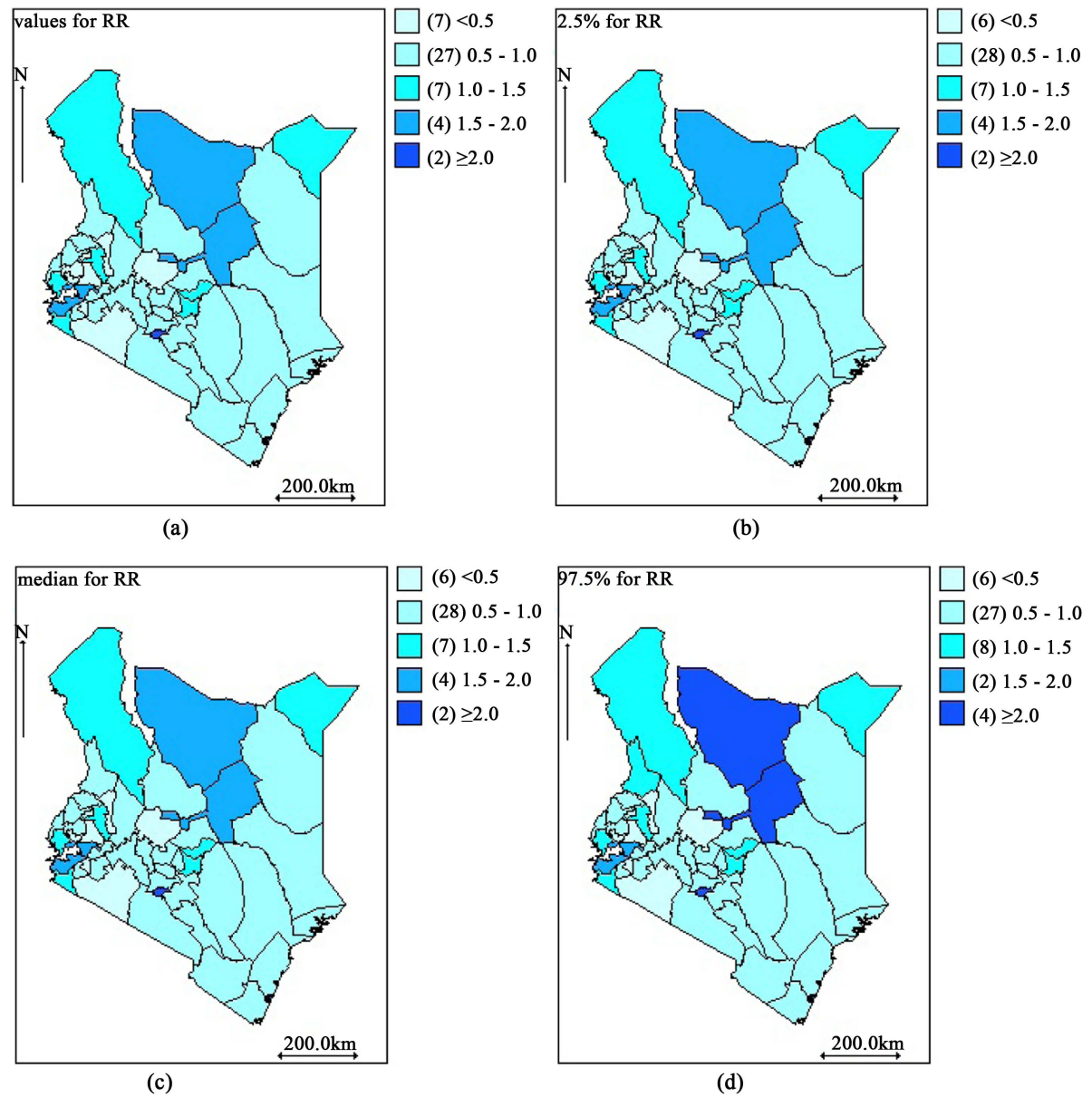
Figure 3. Kenya county level poisson-gamma posterior mean relative risk maps: (a) The mean of the posterior relative risk and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

Gishu, Isiolo, Marsabit, Siaya, Tharaka-Nithi, Mandera, and Embu. This map confirms with **Figure 3** concerning high and low TB prevalence areas.

Despite the fact that assigning a gamma prior distribution for ϑ_i is mathematically convenient, it is likely to be restrictive since covariate adjustment is difficult and there is no possibility for allowing spatial correlation between risk in nearby areas [18]. We therefore present models that nullify theses limitations in the next sections.

5.4.1. Poisson-Log-Normal Model

We now present a model which allows for flexibility of covariates adjustments or incorporation. [18] noted that in disease mapping, the log-normal model is important as it provides a specification that allows for incorporation of covariates. In section we consider LN model without incorporating random effects and covariates and LN model which takes into account random effects and covariates. The random effect term captures or explains
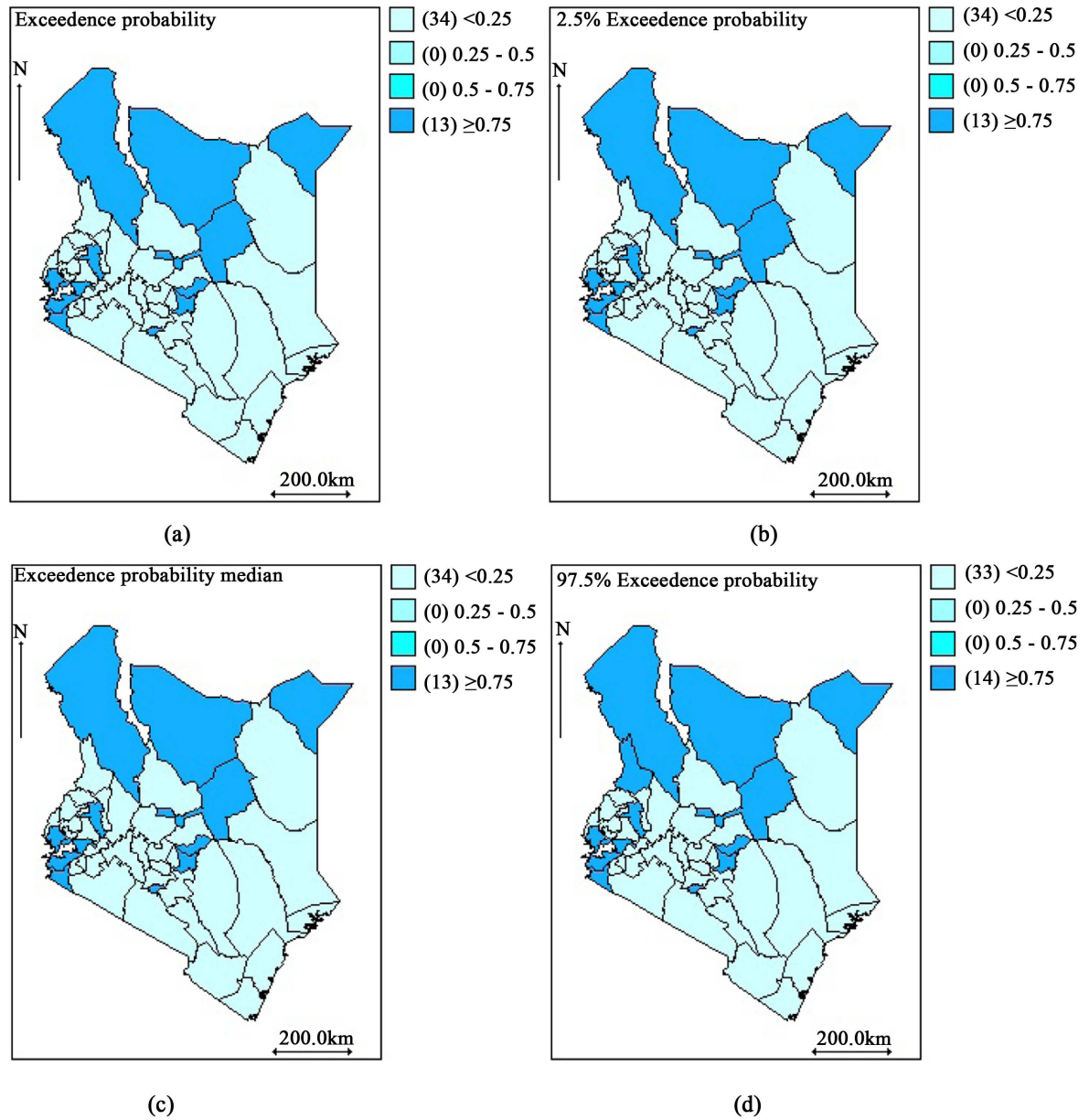
489

Figure 4. Poisson-Gamma posterior relative risk exceedence probability map: Row-wise from the top left figure: (a) the posterior mean relative risk exceedence probability and (b) its 3 2.5% quantile for relative risk, (c) median for the relative risk and (d) 97.5% quantile for the relative risk.

heterogeneity in relative risks of TB among counties. The LN mode with the random effect is often refer to as the unstructured heterogeneity (UH) model.

5.4.2 Model Description

Let y_i and E_i be the observed number and the expected number of disease counts in region $i, i = 1, 2, \cdots, n$ respectively. Further let ϑ_i be the relative risk of disease in region i. We first consider a situation of a Poisson LN model with no area-specific random effect u_i and covariate. As stated in the previous section, $y_i \sim \text{Poisson}(E_i \exp(\eta_i))$, where $\vartheta_i = \exp(\eta_i)$ is the exponential of the linear link function and $\mu_i = E_i \exp(\eta_i)$ is the Poisson mean. Fitting a generalized linear model with a log-link function, we have $\log(\mu_i) = \log(E_i) + \eta_i$. By Bayesian paradigm, we assumed that $\eta_i \sim N(\mu, \tau^2)$ and its hyper-parameters,

$\mu \sim N\left(0,1 \times 10^{-6}\right)$ and $\tau^2 \sim \text{Gamma}\left(0.5, 0.05\right)$. Parameter estimation was carried out using Bayesin Markov Chain Monte Carlo via Gibbs Sampling. Convergence of the MCMC was reached at 11,000 iteration after a burnin period of 10,000 sample and thinning of every 30th element of the chain. Convergence diagnosis plots are presented in **Figure 5** and posterior statistics of parameters are presented in **Table 3**.

We now consider a Poisson LN model with area-specific random effect or uncorrelated heterogeneity (UH) effect u_i and c covariate(s) for region i denoted by X_{ic}. Let $\boldsymbol{X}$ represents the covariates matrix. The Poisson LN non-spatial model is given by

$$y_i \mid E_i, X_{ic}, \eta_i \overset{ind}{\sim} \text{Poisson}\left(E_i \exp\left(\eta_i\right)\right), \tag{12}$$

where $\eta_i = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ic} + u_i$ is the linear link function, u_i are the residual random effects that capture the residual unexplained log relative risk in region i and τ_u^2 is the precision variance. This implies that

$$\log\left(\vartheta_i\right) = \eta_i = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ic} + u_i. \tag{13}$$

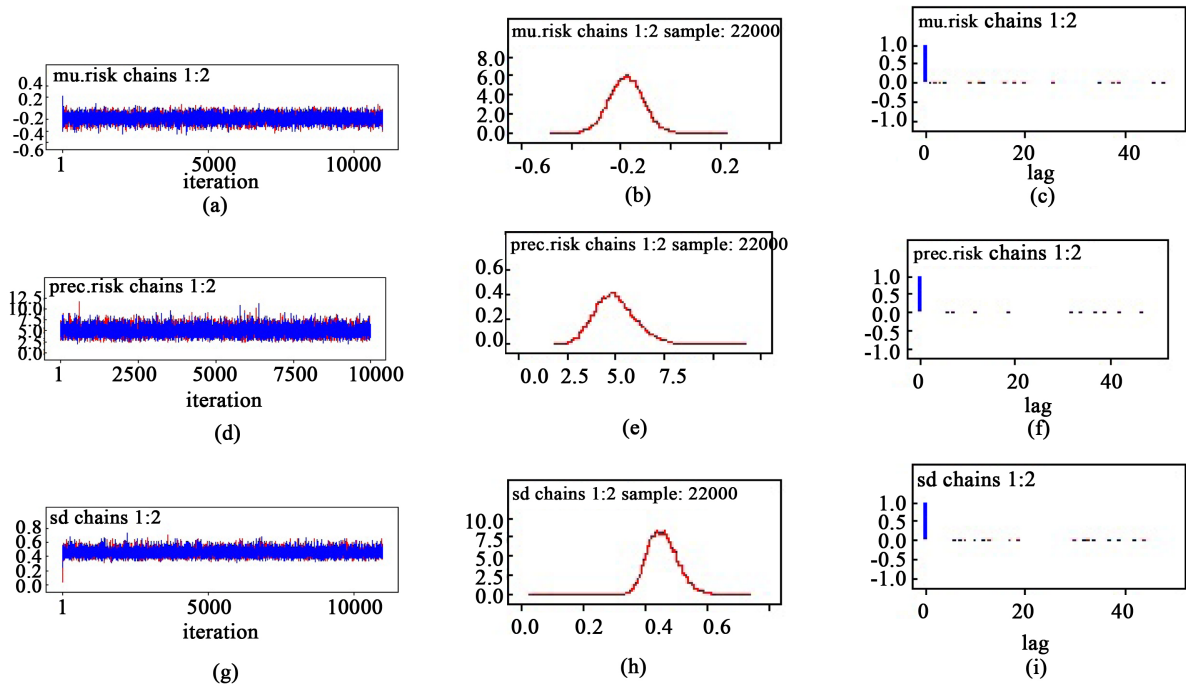From Equation (13), we can write the relative risk as

Figure 5. Poisson log-normal model: convergence diagnosis of Markov Chain Monte Carlo.

Table 3. Posterior statistics of Poisson log-normal model without covariate and random effect.

Parameters	Posterior Means	Credible Region
μ	−0.17	(−0.3093, −0.04605)
τ^2	5.012	(3.182, 7.235)
σ	0.4558	(0.3719, 0.5608)
$\bar{D}$	575,821	-
pD	47.013	-
DIC	622.834	-

$$\vartheta_i = \exp\left(\beta_0 + \sum_{p=1}^{P} \beta_p X_{ic}\right), \tag{14}$$

where ϑ_i are the relative risk of region i, $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)'$ are regression parameters and β_0 is the intercept or the overall risk effect. Here, the mean μ_i of the Poisson distribution is

$\mu_i = E_i \exp(\eta_i) = E_i \exp\left(\beta_0 + \sum_{p=1}^{P} \beta_p X_{ic}\right)$. Fitting a generalized linear model with a log-link function, we have

$$\log(\mu_i) = \log(E_i) + \beta_0 + \sum_{p=1}^{P} \beta_p X_{ic}. \tag{15}$$

5.4.3. Parameter Estimation

Since $y_i \sim \text{Poisson}(E_i \exp(\eta_i))$, the likelihood function of y_i is defined by

$$\ell(\boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{u}) = \prod_{i=1}^{n} \frac{(E_i \exp(\eta_i))^{y_i} \exp(-E_i \exp(\eta_i))}{y_i!} = P(\boldsymbol{y}, \boldsymbol{E} \mid \boldsymbol{\beta}, \boldsymbol{\vartheta}, \boldsymbol{u}), i = 1, 2, \cdots, n. \tag{16}$$

We would need the prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{u}$ to obtain the posterior distribution for the parameters of interest. It is assumed that $\boldsymbol{u} \sim N(0, \sigma^2)$ with mean zero and variance σ^2. We assume a conjugate prior of the Gaussian distribution for the parameter $\boldsymbol{\beta}$ defined as

$$p(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\beta}' \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta}\right) \tag{17}$$

We now show that the posterior distribution of $\boldsymbol{\beta}$ has the Gaussian distribution Section 5.2.

5.5. The Likelihood Function of a Regression with Gaussian Random Effect

Consider a response random variable $\boldsymbol{y} \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix. The linear regression function of y_i on X_{ic} is defined by $y_i = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ic} + u_i$. Therefore, the Gaussian process of regression density or likelihood function for y_i is given by

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right) = N(0, \boldsymbol{\Sigma}), \tag{18}$$

where $\boldsymbol{X}$ is a design matrix of the covariates. Just to simplify analytic calculation, we can alternatively write the Gaussian linear model Equation (18) as

$$P(\boldsymbol{y} \mid \boldsymbol{\beta}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{M}\boldsymbol{\beta})' (\boldsymbol{z} - \boldsymbol{M}\boldsymbol{\beta})\right], \tag{19}$$

where $z_i = \dfrac{y_i}{\sigma_i^2}$, $M_{ci} = \dfrac{X_{ci}}{\sigma_i^2}$, and σ_i^2 are the diagonal covariance matrix $\boldsymbol{\Sigma} = \{\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\}$ with the standard model: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$. Consider a general likelihood function, $\ell(\boldsymbol{\beta})$ and let us take a second order Taylor expansion of the log-likelihood $\ln \ell(\boldsymbol{\beta})$ around its maximum, then we have

$$\ln \ell(\boldsymbol{\beta}) = \ln \ell(\boldsymbol{\beta}_{ML}) + \frac{\partial \ln \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}_{ML}} (\boldsymbol{\beta} - \boldsymbol{\beta}_{ML}) + \frac{1}{2} \frac{\partial^2 \ln \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}\bigg|_{\boldsymbol{\beta}_{ML}} (\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})^2, \tag{20}$$

where $\boldsymbol{\beta}_{ML}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$. Letting $\Omega = \dfrac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}|^{1/2}}$ and taking log of Eq-

uation (19), we have

$$\ln P(y \mid \beta) = \ln \Omega - \frac{1}{2}(z'z - 2z'M\beta + \beta'M'M\beta) \tag{21}$$

Finding the derivative of Equation (21), it follows that

$$\frac{\partial \ln P(y \mid \beta)}{\partial \beta} = z'M - \beta'M'M = 0 \tag{22}$$

Solving Equation (22), we have

$$\beta_{ML} = (M'M)^{-1} M'z. \tag{23}$$

Now finding the second derivative of Equation (25), we have:

$$\frac{\partial^2 \ln P(y \mid \beta)}{\partial \beta^2} = -\frac{\partial(\beta'M'M)}{\partial \beta} = -M'M = H. \tag{24}$$

Hence, we can rewrite Equation (22) as

$$\beta_{ML} = (M'M)^{-1} M'z = H^{-1}M'z. \tag{25}$$

From the Taylor's expansion in Equation (20), and by the Maximum Likelihood Principle (MLP) that

$$(\beta - \beta_{ML}) \frac{\partial \ln P(y \mid \beta)}{\partial \beta} \bigg|_{\beta_{ML}} = 0,$$

it follows that

$$\ln P(y \mid \beta) = \ln P(y \mid \beta_M) - \frac{1}{2}(\beta - \beta_{ML})' H (\beta - \beta_{ML}) \big|_{\beta_{ML}} \tag{26}$$

Taking exponent on both side of Equation (26), we have

$$P(y \mid \beta) = P(y \mid \beta_{ML}) \exp\left[-\frac{1}{2}(\beta - \beta_{ML})' H (\beta - \beta_{ML})\right] \tag{27}$$

or

$$P(y \mid \beta) = \ell_{ML} \exp\left[\frac{1}{2}(\beta - \beta_{ML})' H (\beta - \beta_{ML})\right], \tag{28}$$

where

$$\ell_{ML} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(z - M\beta_{ML})' (z - M\beta_{ML})\right]. \tag{29}$$

5.6. Posterior Function of Gaussian Process Regression

Assuming that the prior distribution of β is

$$p(\beta) \propto \exp\left(-\frac{1}{2}\beta'\Sigma_\beta^{-1}\beta\right). \tag{30}$$

Then writing only the terms from the likelihood and prior which depend on the weights, and "completing the squares" for Multiple parameters model, the posterior function is defined as

$$P(\beta \mid X, y) \propto \exp\left[-\frac{1}{2}(y - X'\beta)' \Sigma^{-1} (y - X'\Sigma)\right] \exp\left(-\frac{1}{2}\beta'\Sigma_\beta^{-1}\beta\right) \tag{31}$$

Simplifying Equation (31), it follows that

$$P(\beta \mid X, y) \propto \exp\left[-\frac{1}{2}(y'\nu y - 2\beta'X'\nu y + \beta'(X'\nu X + \kappa I)\beta)\right], \tag{32}$$

where $v = \boldsymbol{\Sigma}^{-1}$ and $\kappa = \boldsymbol{\Sigma}_p^{-1}$ are the covariance of the likelihood function and the prior function respectively. The Equation (32) above is indeed Gaussian with the constant term $\boldsymbol{y}'\boldsymbol{y}$. In "completing the squares", we are given a quadratic form defining the exponent terms in a Gaussian distribution, and we need to determine the corresponding mean and covariance. To avoid computational complexity with "completing of squares", we sort to using "kernel's trick" TB324. The exponent of a general Gaussian distribution defined as $(\boldsymbol{y} - \mu)' \boldsymbol{\Lambda}(\boldsymbol{y} - \mu)$, where $\boldsymbol{\Lambda}$ is the precision matrix can be expressed as

$$(\boldsymbol{y} - \mu)^t \boldsymbol{\Lambda}(\boldsymbol{y} - \mu) = \left[\boldsymbol{y}'\boldsymbol{\Lambda}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{\Lambda}\mu + \mu'\boldsymbol{\Lambda}\mu\right] = \boldsymbol{y}'\boldsymbol{\Lambda}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{\Lambda}\mu + \text{constant}. \tag{33}$$

Comparing Equation (32) with Equation (33), we have

$$\boldsymbol{\Lambda} = \boldsymbol{X}'v\boldsymbol{X} + \kappa\boldsymbol{I}, \quad \mu = v\boldsymbol{\Lambda}^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{34}$$

That is $P(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}) \sim N(\boldsymbol{\beta} \mid \mu, \boldsymbol{\Lambda}^{-1})$. Note that $\boldsymbol{\Lambda}^{-1}$ must be invertible, that is $|\boldsymbol{\Lambda}| \neq 0$. The maximiser of the likelihood is the mean μ which is again the mode of the likelihood. Therefore, the Maximum Posterior (MAP) is given by

$$\text{MAP} = v(v\boldsymbol{X}'\boldsymbol{X} + \kappa\boldsymbol{I})^{-1} \boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{I})^{-1} \boldsymbol{X}'\boldsymbol{y}. \tag{35}$$

In fact the MAP is similar to the maximum likelihood value $\boldsymbol{\beta}_{ML} = (\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{z}$. Therefore, the posterior mean and Covariance are respectively defined by

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{I})^{-1} \boldsymbol{X}'\boldsymbol{y} \quad \text{and} \quad \boldsymbol{\Lambda}^{-1} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{I})^{-1} \tag{36}$$

It follows that the posterior distribution for $\boldsymbol{\beta}$ is also Gaussian defined by

$$P(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right). \tag{37}$$

Therefore, the prior distribution of $\boldsymbol{\beta}$ is assumed to be normally distributed as

$$P(\boldsymbol{\beta}) = \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^P \exp\left(-\frac{1}{2}\sum_{p=0}^P \frac{\beta_p^2}{\tau_\beta^2}\right) \tag{38}$$

and the prior distribution for the area-specific random effect defined by

$$P(\boldsymbol{u}) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\tau_u}\right)^n \exp\left(-\sum_{i=1}^n \frac{u_i^2}{2\tau_u^2}\right). \tag{39}$$

Therefore, the posterior distribution is defined as

$$P(\boldsymbol{\beta}, \boldsymbol{u}, \tau_\beta^2, \tau_u^2 \mid \boldsymbol{y}, \boldsymbol{E}, \vartheta) \propto P(\boldsymbol{y}, \boldsymbol{E} \mid \vartheta, \boldsymbol{\beta}, \boldsymbol{u}) P(\boldsymbol{\beta}) P(\boldsymbol{u}) \tag{40}$$

Hence,

$$P(\boldsymbol{\beta}, \boldsymbol{u}, \tau_\beta^2, \tau_u^2 \mid \boldsymbol{y}, \boldsymbol{E}, \vartheta) = \prod_{i=1}^n \frac{(E_i \exp(\eta_i))^{y_i} \exp(-E_i \exp(\eta_i))}{y_i!}$$
$$\times \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta^2}\right)^P \exp\left(-\frac{1}{2}\sum_{p=0}^P \frac{\beta_p^2}{\tau_\beta^2}\right) \times \left(\frac{\tau_u^2}{2\pi}\right)^{1/2} \exp\left(-\sum_{i=1}^n \frac{u_i^2}{2\tau_u^2}\right). \tag{41}$$

The prior distribution for the linear regression coefficients is given by $\boldsymbol{\beta} \sim N(0, \tau_\beta^2)$. The corresponding conjugate prior distribution for τ_β^2 is the inverse-gamma [16] [19] defined as

$$P(\tau_\beta^2 \mid \omega, \phi) = \frac{\phi^\omega}{\Gamma(\omega)} (\tau_\beta^2)^{-(\omega+1)} \exp\left(-\frac{\phi}{\tau_\beta^2}\right), \tau_\beta^2, \omega, \phi > 0. \tag{42}$$

The Equation (42) is the hyperprior distribution for τ_β^2 with hyper-parameters (ω, ϕ). We defined $\tau_\beta^2 \sim \text{Gamma}(0.05, 0.005)$ and modeled the random effect $u_i \sim N(0, \tau_u^2)$ and hyper-prior distribution for the

precision parameter $\tau_u^2 \sim \text{Gamma}(0.05, 0.005)$.

We first consider the LN without covariates. Parameter estimation was carried out using Bayesian Markov Chain Monte Carlo via Gibbs Sampling. MCMC convergence was reached at 100,000 iterations after a burnin period of 10,000 sample and thinning of every 30th element in the sample. **Figure 5** presents convergence diagnostics plots of this model. **Table 3** presents the Posterior statistics of the LN model.

Table 3, revealed that the overall mean of the posterior relative risk is −0.17 (95% credible interval = (−0.31, −0.046). This indicates that the overall TB risk effect in Kenya estimated by the Poisson Log-Normal model decreases keeping all other determinants of TB constants. The standard deviation of the relative risk is 0.46 (95% credible interval = 0.37 - 0.56) with precision variation $\tau^2 = 5.01$ (95% credible interval = 3.18 - 7.24) indicating significance of variability of TB risk among counties. In a situation of rare TB cases, standard deviation of the Log-Normal model is expected to be much lower than that of the standardized mortality ratio 0.49 [18].

Figure 6, revealed that TB risk is expected to be high in the North, West, North-West and central counties of
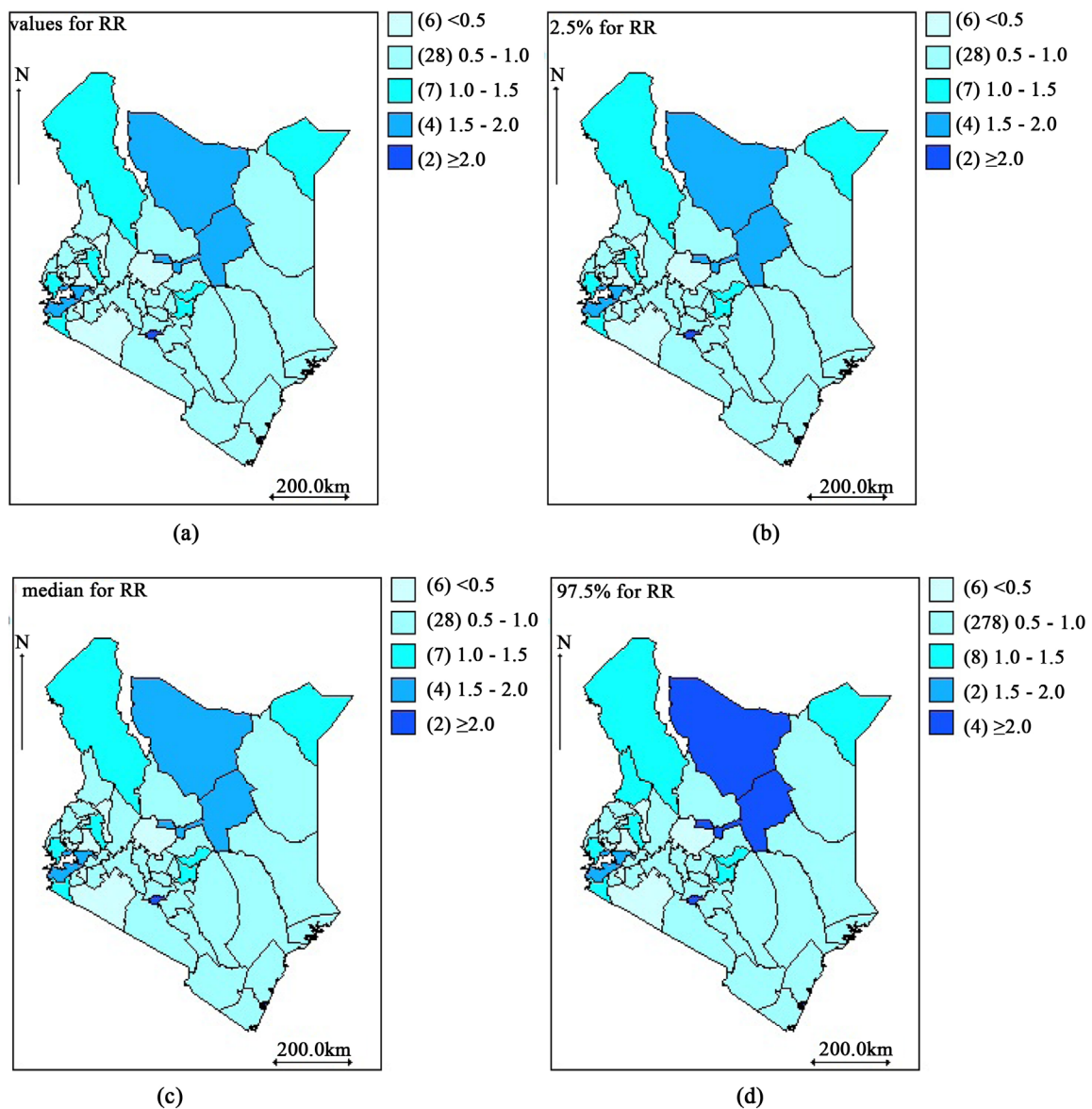

Figure 6. Kenya County level TB prevalence counts: Poisson log-normal model without covariate and random effect. (a) the posterior relative risk map and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

Kenya and low risk in the South-West counties. According to this model, Nairobi and Mombasa are expected to have the highest TB risk (RR > 2.0) and Nandi, Narok, Nyamira, Vihiga, and laikipia are expected to have the lowest TB risk (RR < 0.5). Counties with relative risk above the national risk (RR = 1) are apparent from **Figure 7**. Again, due to abundant of data or information, the range of the posterior relative risk of the Poisson Log-Normal remains the same to that of the SMR. The lowest estimated risk is 0.40 and highest estimated risk is 2.38. There is no reduction of relative risk range compare to SMR's risk as would be expected in a case of rare information or data.

The exceedence probability map **Figure 7** confirmed with the Poisson-Gamma model that 13 counties have their TB risk above the national risk. These counties are: Nairobi, Mombasa, Kisumu, Turkan, Migori, Homa bay, Uasin Gishu, Isiolo, Marsabit, Siaya, Tharaka-Nithi, Mandera, and Embu. **Figure 7** confirms with **Figure 6** that high risk of TB prevalence is observed in the North, West, North-West and central counties and low risk in the South-East counties except Mombasa.
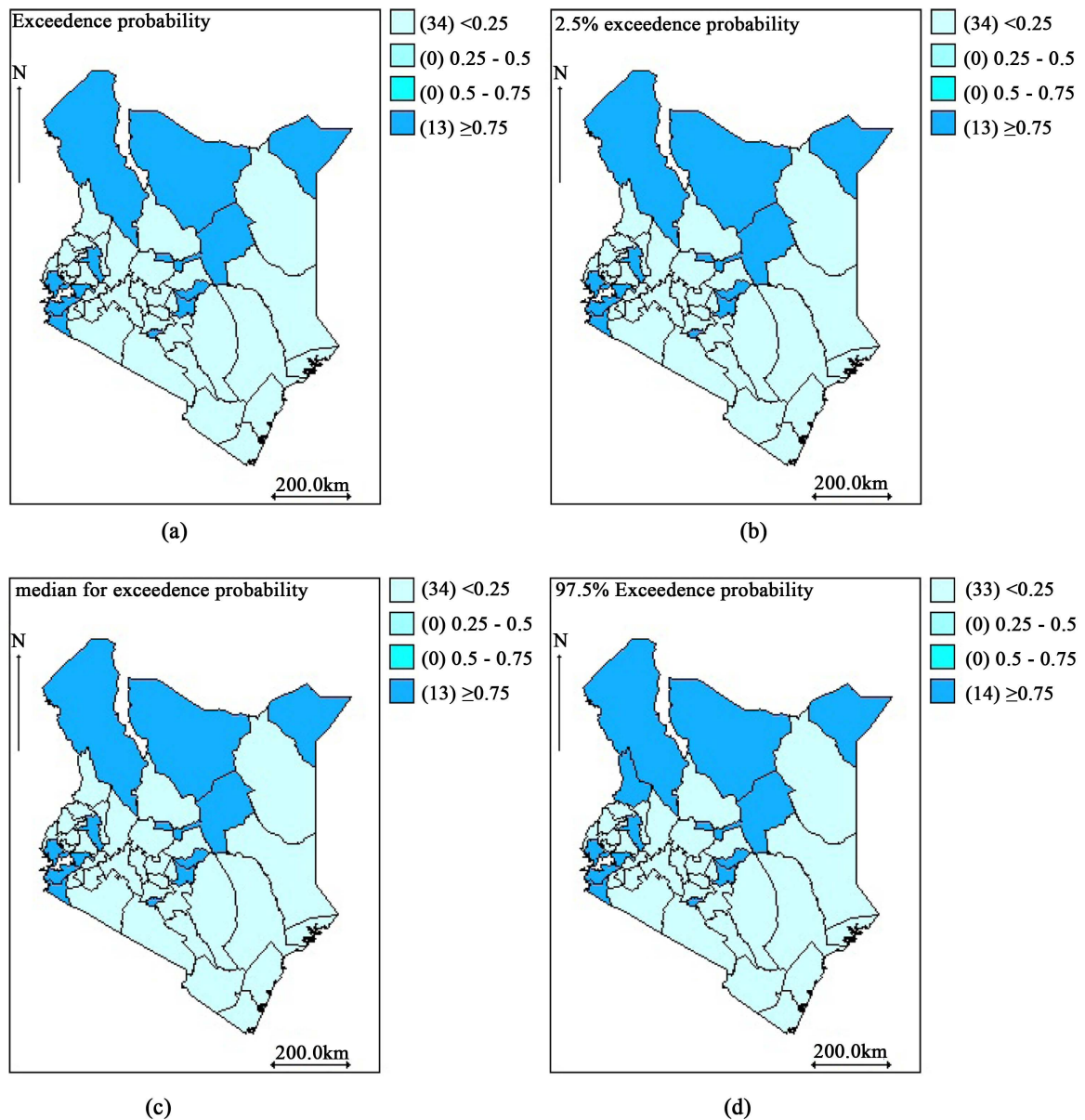

Figure 7. Kenya County level TB prevalence counts: poisson log-normal model without covariate and random effect. (a) the posterior relative risk exceedence probability map and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

We now present the results of the Poisson log-normal model with with UH and covariates effects. **Figure 8** presents the Rubin and Gelman convergence diagnostics plots of this mode. **Table 4** revealed that the overall level of relative risk effect estimated is $\beta_0 = -0.18$ (95% credible interval = (−0.30, −0.06)). The overall risk effect is significantly different from zero and negative. This indicates that overall TB risk effect would be decreasing keeping all other determinants of TB constants. Among the cova- riates considered as TB determinants, only HIV parameters is significantly different from zero 1.20 (95% credible interval = 0.50 - 2.60) but positive. This indicates that TB risk increases with increasing HIV prevalence. [18] noted that, the higher the τ_u^2, the higher the variability of TB risk among counties and and the lower it is the lower the variability. This means that, a very small τ_u^2 will indicate possibility of risk similarity between neighbouring counties. The precision for the UH $\tau_u^2 = 5.32$ (95% credible interval = 3.35 - 7.76) is significant, indicating that there exist variation in risk among counties. The UH revealed high variability of relative risk compare to the Poisson Log-Normal without random and covariate effects $\tau^2 = 5.01$ (95% cre- dible interval = 3.18 - 7.24).

Figure 9 also confirmed that out of the 47 counties in Kenya, 13 exhibit TB relative risk higher than the national risk (RR = 1). **Table 5** presents counties in groups according to their relative risk level. High TB risk can again be observed that in the North, West, North-West and central counties of Kenya and low TB risk in the South-East counties except Mombasa.

Table 5 presents the results of the UH model with counties categorised according to their range of relative risk. The results showed that 14 counties have their relative risk above 1 and the lowest risk counties are 5. The exceedence probability map in **Figure 10** visually presents counties with risk above 1. **Figure 10** shows 14 counties having elevated risk of TB. These maps again confirmed high TB risk in the North, West, North-West and central counties of Kenya and low risk in South-East counties of Kenya except Mombasa.

Figure 11 is the maps of the area-specific random effect (u_i), which shows variation of risk among counties in Kenya. This map captures and displays true TB excess risk surface after covariates and confounding factors are considered TB21. Excess risk of TB is observed in Marsbit, Embu, Migori and Kisumu.

Table 4. Posterior statistics of the poisson log-normal model with UH effect.

Model indicators	UH
β_0	−0.1765 (−0.2957, −0.05936)
HIV	1.198 (0.4928, 2.571)
Firewood	0.2735 (−2.215, 2.144)
five kilometer distance	−1.317 (−3.423, 1.437)
σ_u	0.4409 (0.359, 0.5466)
τ_u^2	5.324 (3.347, 7.757)
$\bar{D}$	575,779
pD	46.973
DIC	622.753

Table 5. UH results indicating counties with high and low TB risk.

| RR > 2.0	RR: 1.5 - 2.0	RR: 1.0 - 1.5	RR < 0.5
Nairobi, 2.159 (2.145, 2.174)	Homa bay ,1.721 (1.702, 1.74)	Embu, 1.199 (1.18, 1.219)	Lakaipia, 0.458 (0.4449, 0.4714)
Mombasa, 2.383 (2.36, 2.407)	Isiolo, 1.957 (1.906, 2.01)	Mandera, 1.044 (1.02, 1.067)	Nandi, 0.4033 (0.3941, 0.4127)
-	Kisumu, 1.975 (1.955, 1.995)	Migori, 1.374 (1.357, 1.392)	Narok, 0.482 (0.4715, 0.4928)
-	Marsabit, 1.969 (1.93, 2.008)	Siaya, 1.401 (1.384, 1.419)	Nyamira (Kisii North), 0.4533 (0.4422, 0.4646)
-	-	Tharaka-Nithi, 1.064 (1.042, 1.086)	Vihiga, 0.469 (0.4581, 0.4801)
-	-	Turkan, 1.068 (1.051, 1.086)	-
-	-	Uasin Gishu, 1.182 (1.166, 1.198)	-
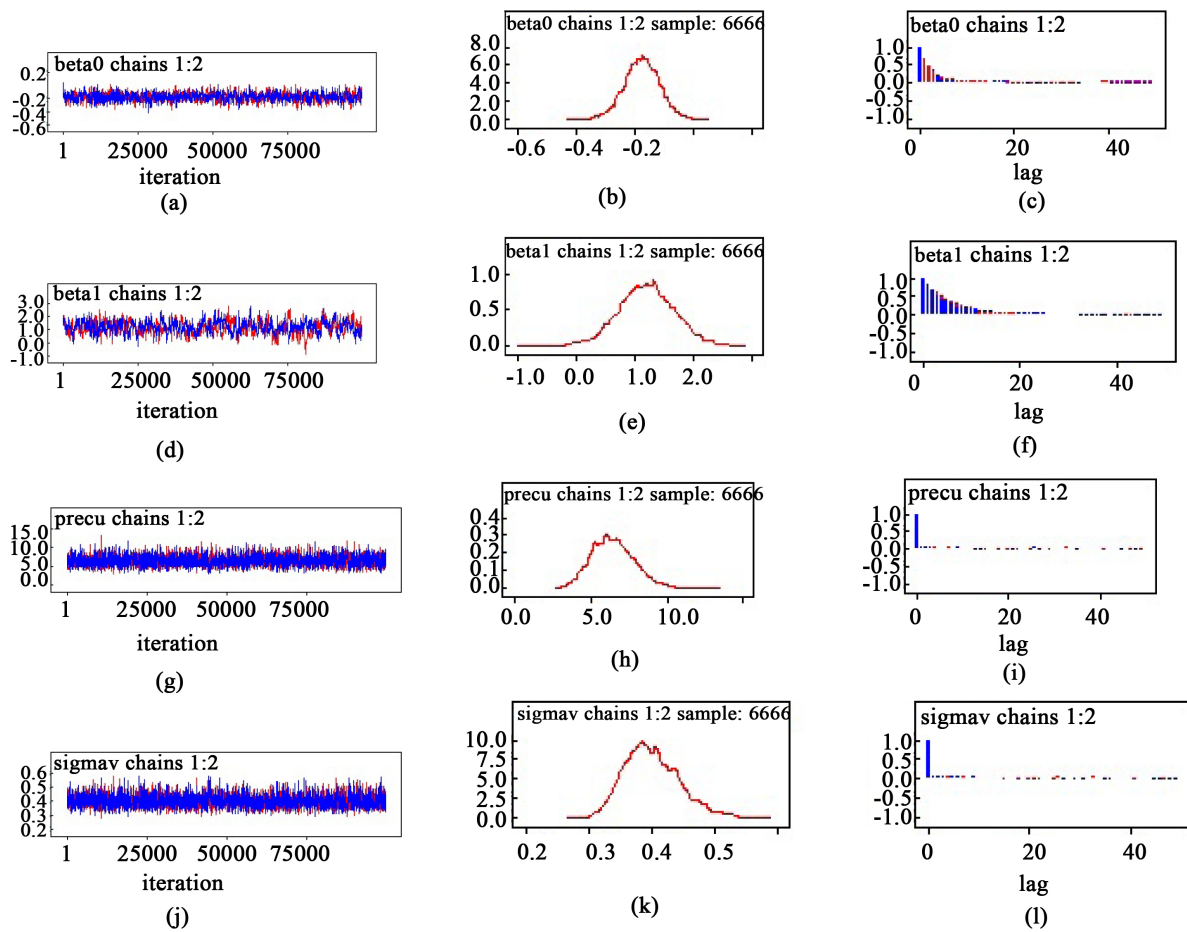
Figure 8. Poisson log-normal model with UH and covariates effects: convergence diagnostics of Markov Chain Monte Carlo.

5.7. Summary of the Non-Spatial Models

Though the Poisson-Gamma model provides good information about the TB prevalence in Kenya, one of its shortcomings is that it is unable to handle problem of spatial correlation and incorporation of covariates [18]. The Poisson Log-Normal model provides specifications that allow for incorporation of covariates. The Poisson Log-Normal model also enable us to capture the area random effect and to explain the extend of risk variability among counties through the unstructured random effect term u_i.

Though thinning reduces the speed of the MC but it significantly reduces the number of iterations and solves the issue of autocorrelation among parameters that form the chain. Thinning reduces storage demand while preserving the integrity of the MC process [19]. [19] noted that the value of every k^{th} element to be sampled is determined by the researcher and out most care must be taken since extremely large k value may results in lost of potentially an important information.

The slower the chain to converge, the more careful one should be about the burn-in period. However, it should be noted that there is no standard, systematic or guaranteed way of determining the length of the burn-in period [19]. Nevertheless, considerable work on convergence diagnostics has been done to make specific recommendations and identify tests [19].

HIV is identified as significant among the covariates considered. Reason being that HIV patients have their immune system weakened or destroyed by the HIV rendering the body natural defence incapable of carrying out its function of protecting the body against other diseases. Since HIV has this capacity to weaken the immune system, it also implies that it has the effect of re-activating latent TB to active TB in individuals who are latently infected. Models that allow for handling spatial correlation are discussed in Chapter 6.
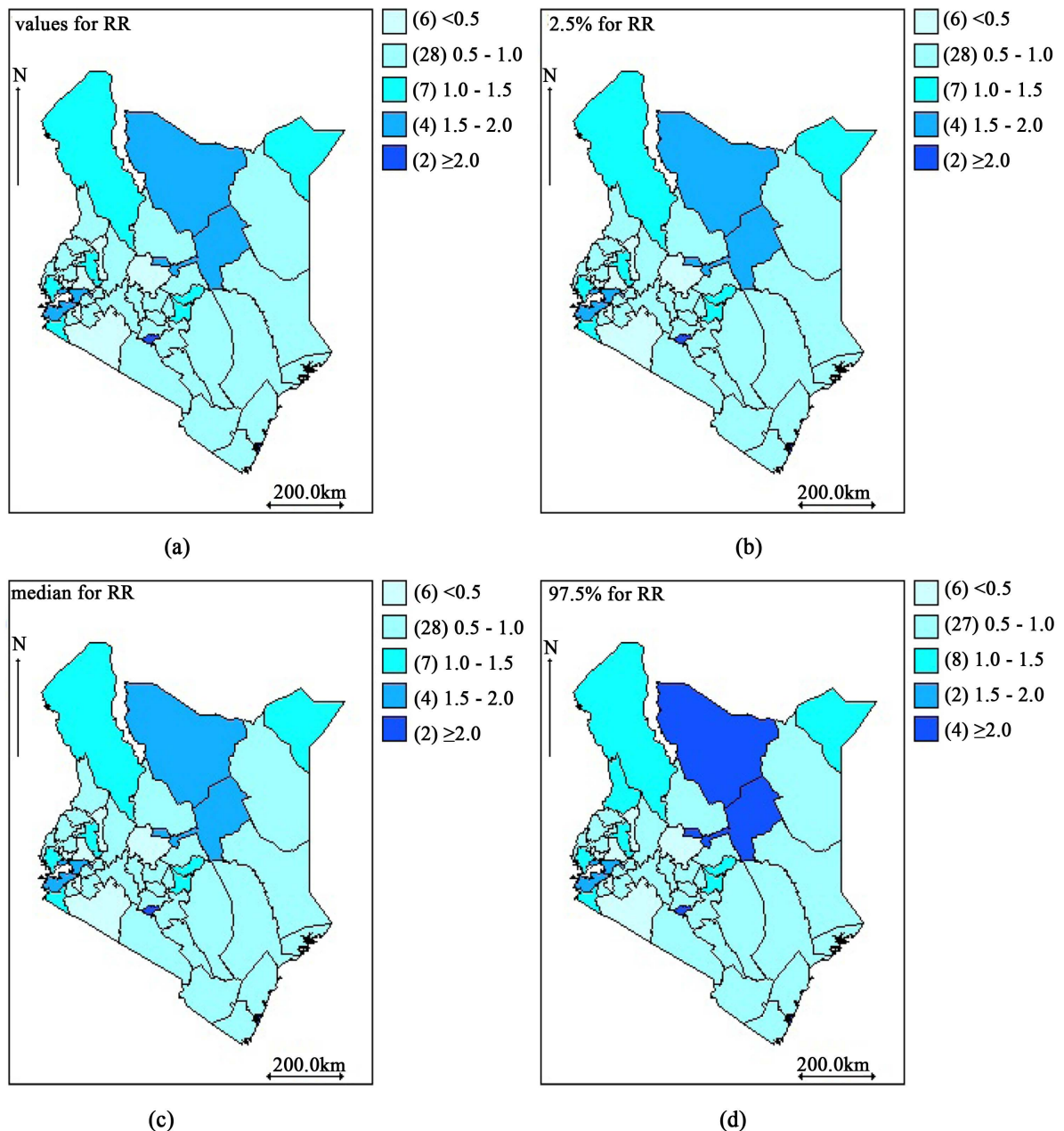
Figure 9. Kenya county level TB prevalence counts: UH smooth relative risk map (a) and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

6. Bayesian Hierarchical Spatial Model for Disease Mapping

This section presents spatial models used to identify and detect clustering of disease risk in the study area of interest. Spatial data are directly or indirectly referenced to a location on the surface of the earth. These models would allow for borrowing of strength between neighbouring counties such that neighbouring counties shall have similar risk whiles distant counties are expected to show variation in risk. The idea of spatial auto-correlation in spatial data analysis is that values of variables in near-by locations are more similar or related than those far apart. Waldo Tobler's first law of spatial analysis states that "everything is related to everything else but near-by things are more related than distant things" [20]. In particular, we investigate the statistical pro-perties of the Conditional Autoregressive (CAR) model and the [11] models.
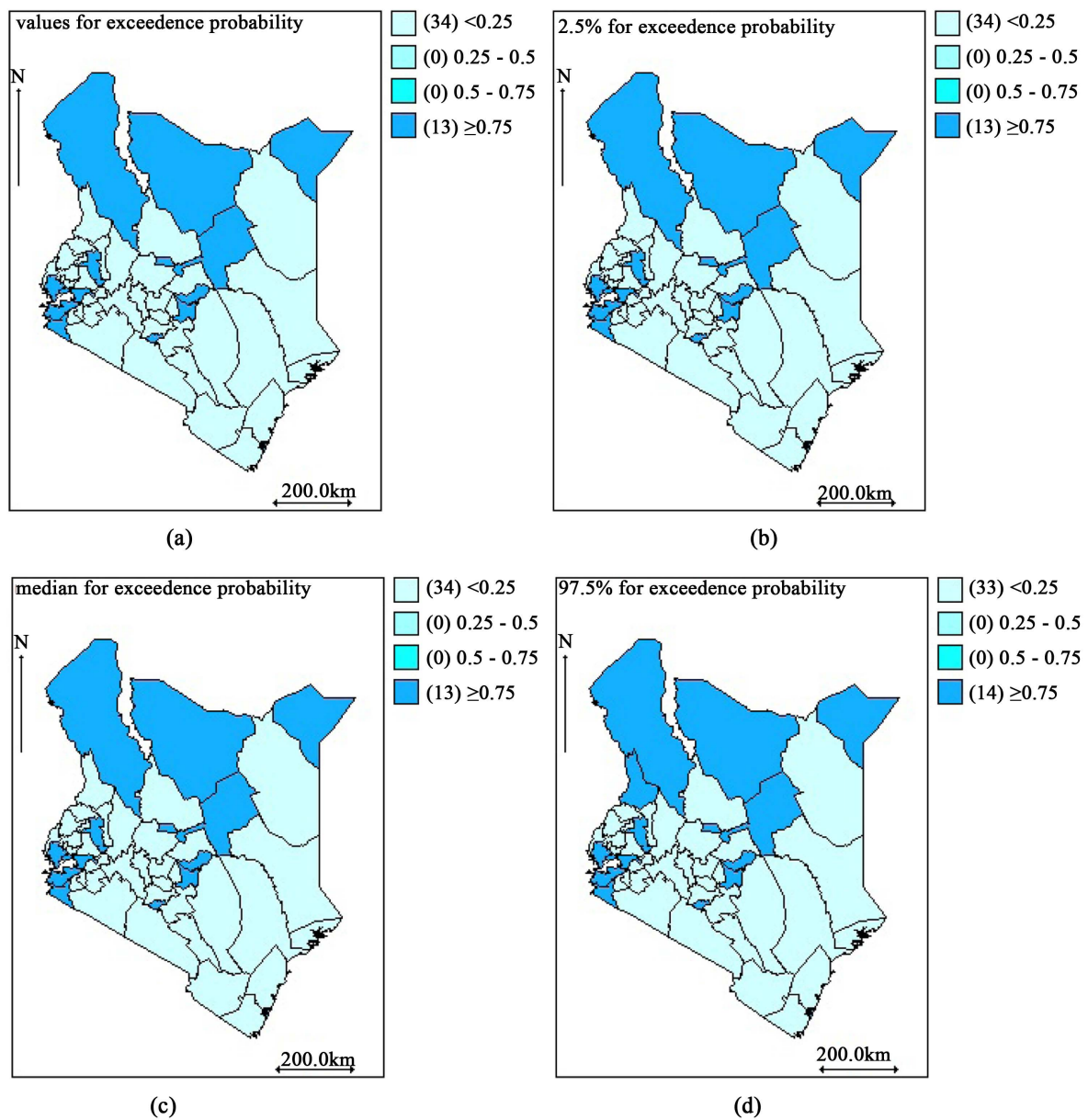
Figure 10. Kenya county level TB prevalence counts: UH smooth relative risk exceedence probability map (a) and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

6.1. Conditional Autoregressive (CAR) Model

Though conditional autoregressive models where introduced decades ago by [11], they were not widely used until the 1990s. Since they are defined conditionally, they are particularly suited for use with the Gibbs sampler [21]. The Conditional Autoregressive (CAR) models have been used extensively to identify and detect clustering in diseases risk. In these models, risks of disease at any given area is affected by the risk in the neighbouring areas. These models have been referred to as the structured model or the Correlated Heterogeneity (CH) models. That is, estimation of risk in any given area depends on risk at neighbouring areas [18]. The distances or boundaries between the regions are often used in determine neighbourhood properties within CAR models [22].

Generally, the CAR model is a continuous Markov random field with a conditional probability density function characterization and designed to model spatial phenomena that are highly related to a specific local context [23] [24]. Application of CAR models can found in [23] [15] [25]. Let $S = \{1, 2, \cdots, n\}$ represents the
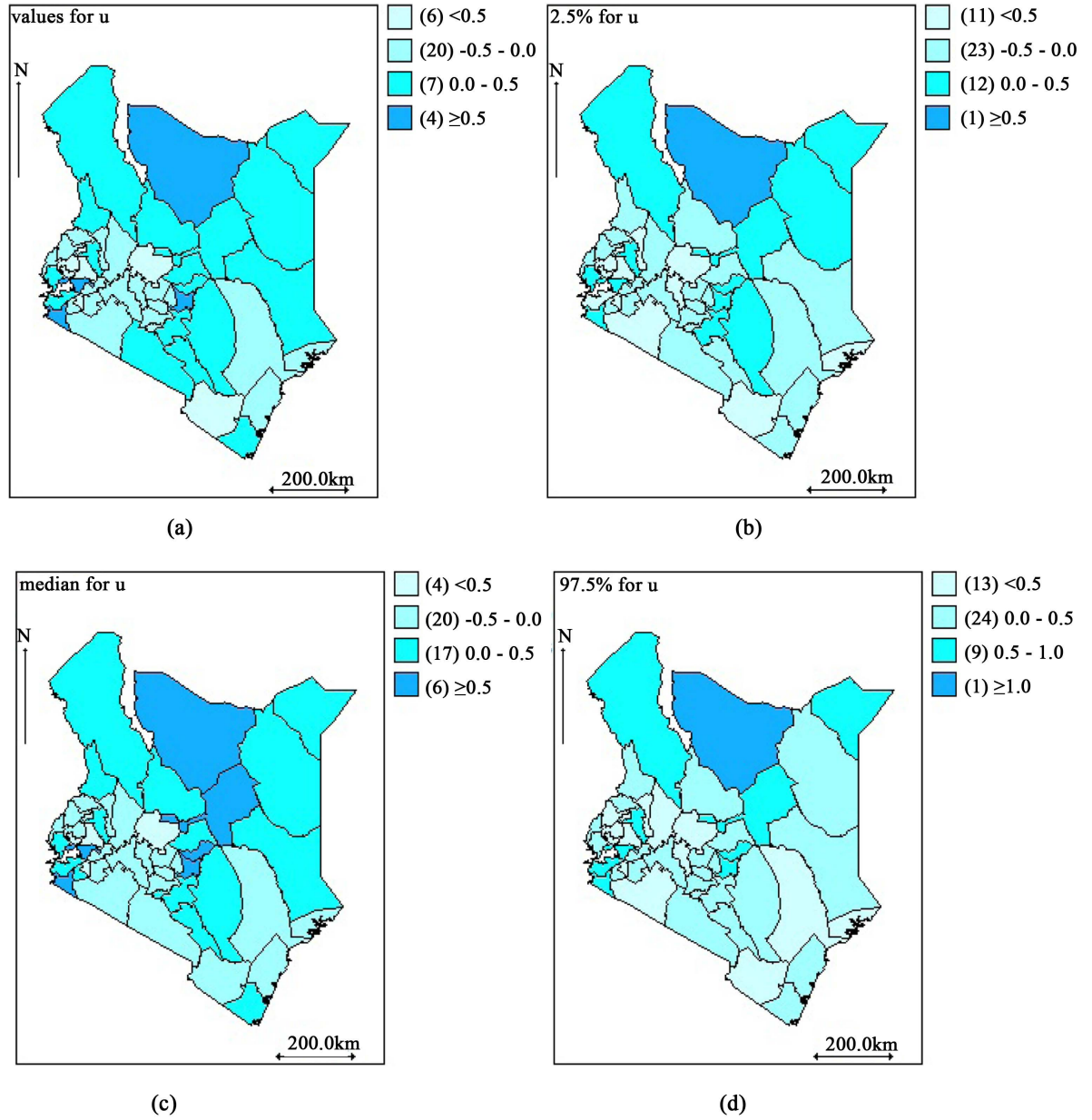
Figure 11. Area-specific random effect: the posterior map (a) and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

area to be studied. Let $N_i = \{ j \in S : i \in j \}$ denote the set of all regions that are neighbouring region i. Let $v_i, i \in S$ be a random variable. We define the corresponding random field v as the vector $v = (v_1, v_2, \cdots, v_n)'$.

In the Gaussian CAR model, we often assume that each observation of the outcome variable v_i has a conditional distribution defined by

$$v_i \mid v_{j \neq i} \sim N \left(\sum_{i \neq j} \Phi_{ij} v_j, \tau_i^2 \right). \tag{43}$$

These are full conditionals where Φ_{ij} is the weight of each observation on the mean of v_i and also denotes the spatial dependence parameter. The Φ_{ij} is non-zero only if $j \in S$. Conventionally, we set $\Phi_{ij} = 0$ since we do not want to regress any observation on itself. Hence no region is a neighbour of itself. The v_j denotes a vector of all observation except v_i. Note that v_i depends only on a set neighbours v_j only if location j is a

neighbourhood set N_i of v_i. The τ_i^2 is a potential unique variance for v_i. For instance, if state i has M neighbours and $\Phi_{ij} = \dfrac{1}{M}$ for every state that is a neighbour, and $\Phi_{ij} = 0$ otherwise, then the conditional expectation of a state's observation is the mean of all neighbours observations [25]. The Gaussian processes are specified by their mean and covariance function [26]. Assuming that each conditional distribution is Gaussian, we will need the mean and the variance-covariance to define the CAR model. The mean and the variance-covariance are respectively defined as

$$E\left[v_i \mid v_{j \neq i}\right] = \mu_i + \sum_{j \in N_i} \Phi_{ij}\left[v_j - \mu_i\right] \quad \text{and} \quad \text{var}\left(v_i \mid v_j\right) = \tau_i^2. \tag{44}$$

Therefore, conditional probability density function of a CAR random variable v_i is has the form [25]

$$f\left(v_i \mid v_{j \neq i} \in S\right) = \sqrt{\frac{1}{2\pi\tau_i^2}} \exp\left\{-\frac{\left[\left(v_i - \mu_i\right) - \rho \sum\limits_{j \in N_i} \Phi_{ij}\left(v_j - \mu_j\right)\right]^2}{2\tau_i^2}\right\}, \tag{45}$$

where $\mu_i = \in \mathbb{R}, \tau_i^2 \in \mathbb{R}^+, |\rho| < 1$, $\Phi_{(ij)} \in \mathbb{R}, \Phi_{(ij)} = \Phi_{(ji)}$, $\Phi_{(ii)} = 0$. The conditional joint probability density function of all the observations is

$$f\left(v_i \mid v_{j \neq i}\right) = \frac{1}{(2\pi)^{n/2} \det\left(B^{-1}\Sigma_D\right)^{1/2}} \exp\left[-\frac{(v - \mu)\Sigma_D^{-1}B(v - \mu)}{2}\right], \tag{46}$$

where $\mu \in \mathbb{R}^{n \times 1}$ (n-dimensional vector), $\mu = \left(\mu_1, \mu_2, \cdots, \mu_n\right)'$, $B \in \mathbb{R}^{n \times n}$ invertible matrix defined as

$$B = (I - \rho\Phi) \quad \text{with} \quad B_{(ij)} = \begin{cases} 1 & \text{if } i = j \\ -\rho\Phi_{(ij)} & \text{if } j \in N_i, \\ 0 & \text{otherwise} \end{cases} \tag{47}$$

$\Sigma_D \in \mathbb{R}^{+n \times n}$ diagonal matrix; $\Sigma_D = diag\left(\tau_1^2, \cdots, \tau_n^2\right) = \tau_i^2$ such that Σ_D is symmetric. It follows that the joint multivariate Gaussian distribution for v_i with $\mu = 0$ has covariance matrix $\Sigma_{D_v} = \Sigma_D^{-1}B = B^{-1}\Sigma_D$ which is symmetric such that $\Phi_{(ij)}\tau_j^2 = \Phi_{(ji)}\tau_i^2, i, j \in S$. Thus, a conditional autoregressive model v in (45) has a probability density function defined as

$$v_i \mid v_{j \neq i} \sim N\left[\mu_i + \rho \sum_{j \in N_i} \Phi_{(ij)}\left(v_j - \mu_j\right), \tau_i^2\right], i \in S \tag{48}$$

and the joint probability density function in (46) becomes

$$v \sim N\left(\mu, B^{-1}\Sigma_D\right). \tag{49}$$

The necessary and sufficient condition for (49) to be a valid joint probability density function is that its covariance matrix should not only be symmetric, but also positive definite (that is, its eigenvalues $\lambda_i > 0, i, \cdots, n$) TB223. For v_i to be a Gaussian random variable, we need to show that Σ_D is symmetric.

To show that Σ_D is symmetric, we defined a symmetric weighted adjacency matrix W, where

$$W = \left(w_{(ij)}\right) \quad \text{with} \quad w_{(ij)} = \begin{cases} 1 & \text{if } j = i \\ \varphi(i, j) & \text{with } j \in N_i : \forall i, j \in S, w_{(ij)} = w_{(ji)} \\ 0 & \text{otherwise,} \end{cases} \tag{50}$$

where $\varphi(i, j)$ is a measure that quantifies the proximity between region i and region j; if $\varphi(i, j) = 1$, then i and j share a common boundary (neighbours). The $\varphi(i, j)$ could be the distance between the centroids of

region i and j. Also, if $\varphi(i, j) = 1$, then j is one of the h nearest neighbours of i. Let W_D be the diagonal of the adjacency matrix W. The adjacency matrix of normalization or standardization W_D is defined as

$$W_D = \text{diag}\left(w_{(1+)}, w_{(2+)}, \cdots, w_{(n+)}\right). \tag{51}$$

Suppose

$$w_{(i+)} = \sum_{j \in N_i} w_{(ij)}, i, j \in S, \tag{52}$$

then we define a matrix of interaction, Θ to be a normalized adjacency matrix defined as

$$\Theta = W_D^{-1} W \quad \text{with} \quad \Theta_{(ij)} = \frac{w_{(ij)}}{w_{(i+)}}, \quad \text{where} \quad \Theta_{(ij)} \tau_j^2 = \Theta_{(ji)} \tau_i^2, i, j \in S. \tag{53}$$

Suppose again that the matrix W_D corresponding to a constant diagonal matrix normalized as (53), then we have
α

$$W_D = \tau^2 W_D^{-1} \quad \text{with} \quad \tau_i^2 = \frac{\tau^2}{w_{(i+)}}, i \in S, \tau^2 \in \mathbb{R}^+. \tag{54}$$

It follows that the conditional joint probability density function can be rewritten as

$$P(v_1, \cdots, v_n) \propto \exp\left\{-\frac{1}{2\tau^2} \Psi' \left(\Sigma_{D_w} - W\right)^{-1} \Sigma_D \Psi\right\}, \tag{55}$$

where $\Psi = (v - \mu)$ and $B = \left(\Sigma_{D_w} - W\right)$. Hence, the CAR model structure for v_i is defined as

$$v_i \mid v_{j \neq i} \sim N\left(\sum_j \frac{w_{ij}}{w_{i+}} v_j, \frac{\tau_v^2}{w_{i+}}\right), \tag{56}$$

and the Equation (49) can be alternatively defined as

$$v \sim \left(\mu, \left[\frac{1}{\tau_v^2}\left(W_D - \rho W\right)\right]^{-1}\right). \tag{57}$$

The τ_v^2 controls the overall variability of v_i, while ρ represents the overall effect of spatial dependence. The value of ρ is should be chosen appropriately [25].

The row stochasticity of $\hat{W} = diag\left(\frac{1}{w_{i+}}\right)W$ indicates that the distribution is improper. This impropriety can be fixed by the parameter ρ. Redefining $\Sigma_{D_v}^{-1} = \left(\Sigma_{D_w} - \rho W\right)^{-1}$ and choose ρ such that $\Sigma_{D_v}^{-1}$ is non singular, preferably with $\rho \in \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_n}\right)$, where $\lambda_1 < \cdots < \lambda_n$ are the ordered eigenvalues of $\Sigma_{D_w}^{-1/2} W \Sigma_{D_w}^{-1/2}$. Simplifying the bounds, we replace W by $\hat{W}$. It follows that

$$\Sigma_{D_v}^{-1} = \Sigma_{D_w}\left(I - \rho \hat{W}\right). \tag{58}$$

If $|\rho| < 1$, then $\Sigma_{D_w}\left(I - \rho \hat{W}\right)$ is non singular. Non-singularity is guaranteed if $\rho \in \left(\frac{1}{\lambda_1}, 1\right)$ where λ_1 is the minimum eigenvalue of $\Sigma_{D_w}^{-1/2} W \Sigma_{D_w}^{-1/2}$. The bound mostly preferred is $\rho \in (0, 1)$. This is a proper Intrinsic Autoregressive model which add parametric flexibility and $\rho = 0$ is an indication of independence. ρ is the additional parameter which makes v_i independent when it is equal to zero. An improper choice of $\rho = 1$ may enable wider scope for posterior spatial pattern and may be preferable [27].

6.2. Parameter Estimation

In this study, we estimate parameters in the CAR model using Bayesian hierarchical methods. In disease mapping, we assumed that disease counts

$$y_i \sim \text{Poisson}\left(E_i \exp(\eta_i)\right), \quad \text{where } \mu_i = E_i \exp(\eta_i) \text{ is the mean of Poisson distribution.} \tag{59}$$

The relative risk is defined as

$$\vartheta_i = \exp(\eta_i), \quad \text{where } \eta_i = X'\boldsymbol{\beta} + v_i, \text{ and } v_i \text{ has a CAR structure.} \tag{60}$$

Fitting the generalized linear model with a log-link function, we have $\log(\mu_i) = \log(E_i) + X'\boldsymbol{\beta} + v_i$. Under the Bayesian method, given the likelihood function of ϑ defined as

$$\ell(\boldsymbol{\beta}, \boldsymbol{v}) = \prod_{i=1}^{n} \frac{\left(E_i \exp(\eta_i)\right)^{y_i} \exp\left(-E_i \exp(\eta_i)\right)}{y_i!} = P(y, E, \vartheta \mid \boldsymbol{\beta}, \boldsymbol{\upsilon}), \tag{61}$$

the prior distribution for $\boldsymbol{\beta}$ is

$$P(\boldsymbol{\beta}) = \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^{P} \exp\left(-\frac{1}{2}\sum_{p=0}^{P} \frac{\beta_p^2}{\tau_\beta^2}\right) \tag{62}$$

and the prior distribution for the CAR random effect is defined by

$$P(\boldsymbol{v}) = \left[v_i \mid v_{j \neq i}, \tau_v^2\right] \sim N\left(\sum_{j \neq i} \frac{w_{ij}}{w_{ij}} v_j, \frac{\tau_v^2}{w_{ij}}\right) \sim CAR\left(0, \tau_v^2\right). \tag{63}$$

The posterior distribution is defined as

$$P\left(\boldsymbol{\beta}, \boldsymbol{v}, \tau_\beta^2, \tau_v^2 \mid y, E, \vartheta\right) \propto P\left(X, E, \vartheta \mid \boldsymbol{\beta}, \boldsymbol{v}, \tau_\beta^2, \tau_v^2\right) P(\boldsymbol{\beta}) P(\boldsymbol{v}). \tag{64}$$

Therefore,

$$P\left(\boldsymbol{\beta}, \boldsymbol{v}, \tau_\beta^2, \tau_v^2 \mid y, E, \vartheta\right) = \prod_{i=1}^{n} \frac{\left(E_i \exp(\eta_i)\right)^{y_i} \exp\left(-E_i \exp(\eta_i)\right)}{y_i!}$$
$$\times \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^{P} \exp\left(-\frac{1}{2}\sum_{p=0}^{P} \frac{\beta_p^2}{\tau_\beta^2}\right) \left(\sum_{j \neq i} \frac{w_{ij}}{w_{ij}} v_j, \frac{\tau_v^2}{w_{ij}}\right). \tag{65}$$

The hyperperprior distribution for the precision parameters τ_v^2 and τ_β^2 are $\tau_v^2 \sim \text{Gamma}(0.05, 0.005)$ and $\tau_\beta^2 \sim \text{Ggamma}(0.5, 0.05)$ respectively. The linear regression coefficient distribution is defined by $\boldsymbol{\beta} \sim N\left(0, \tau_\beta^2\right)$. Parameter estimation was carried out using Bayesin Markov Chain Monte Carlo via Gibbs Sampling. Convergence of the MCMC was reached at 11,000 iteration after a burnin period of 10,000 sample and thinning of every 30th element of the chain. Convergence diagnosis plots are presented in **Figure 12** and posterior statistics of parameters presented in **Table 6**. We compare these results with the [11] model results in Section 6.3.

6.3. The Besag, York and Mollié (BYM) Model

Among the models proposed for performing risk smoothing which have appeared in literature, the [11] model has found most extensive application. The BMY model is divided into two components; the CAR model component v_i, and the UH component, u_i (discussed in Section 6.1 and Section 5.1.1 respectively). The BYM model was introduced by [10] and latter developed by [11]. The BYM or convolution model is defines as

$$\eta_i = \mu + u_i + v_i \tag{66}$$

As noted earlier, we assume

$$y_i \sim \text{Poisson}\left(E_i \exp(\eta_i)\right), \quad \text{where } \mu_i = E_i \exp(\eta_i) \text{ is the mean of the Poisson distribution.} \tag{67}$$

The linear link function $\eta_i = X'\boldsymbol{\beta} + u_i + v_i$. The log relative risk $\log(\vartheta_i) = \eta_i$. Therefore, the relative risk for
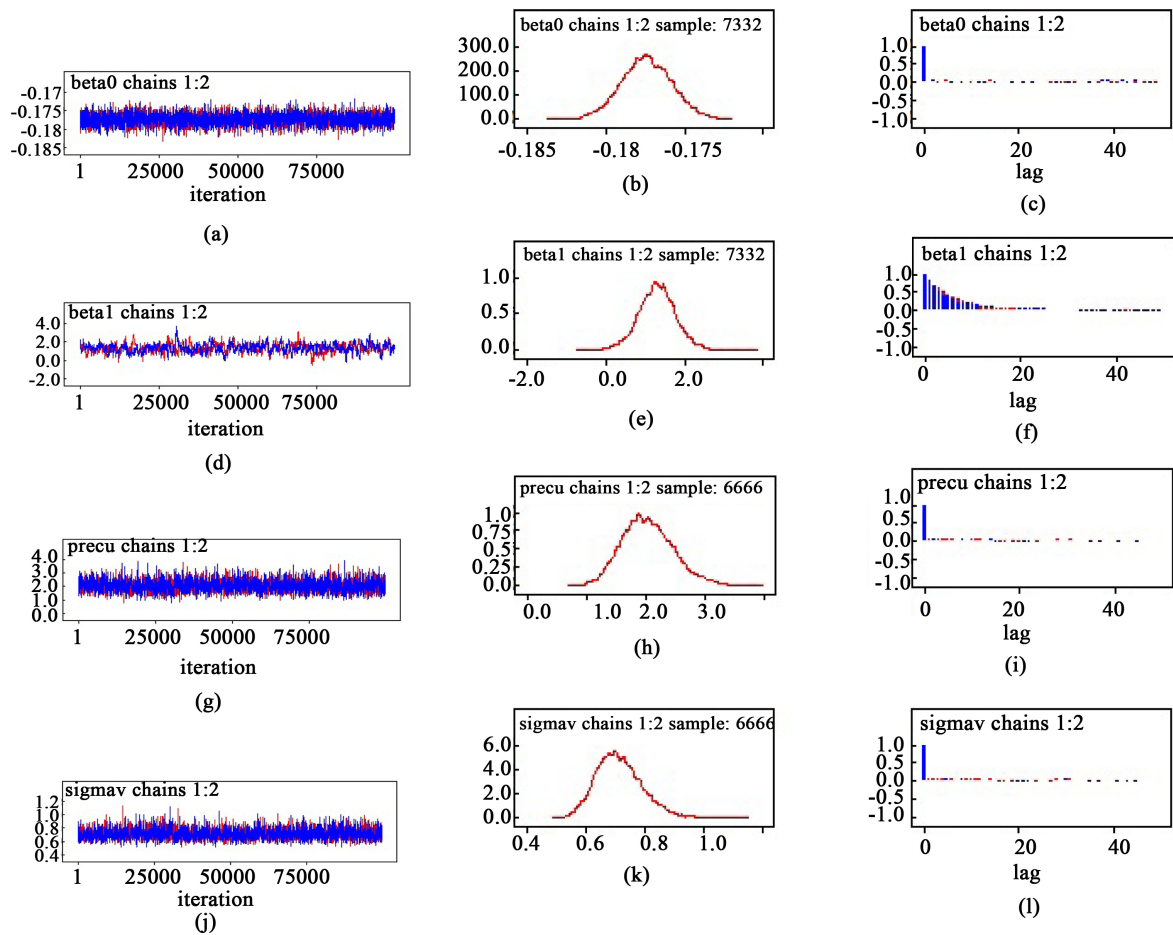
Figure 12. Poisson log-normal model with CAR model: convergence diagnosis of Markov Chain Monte Carlo.

Table 6. Posterior statistics of the CAR and BYM models.

Model indicators	CAR	BYM
β_0	−0.1774 (−0.1805, −0.1743)	−0.179 (−0.267, −0.0908)
HIV	1.812 (0.7735, 2.758)	1.41 (0.488, 2.34)
Firewood	0.2764 (−2.44, 2.822)	−0.28 (−1.29, 0.793)
five kilometer distance	−1.505 (−4.19, 1.18)	−0.852 (−1.81, 0.124)
σ_v	0.8298 (0.6751, 1.03)	0.372 (0.156, 0.678)
τ_v^2	1.559 (0.9432, 2.194)	11.3 (2.18, 40.8)
σ_u	-	0.298 (0.158, 0.416)
τ_u^2	-	13.4 (5.79, 40)
$\bar{D}$	578,018	50.969
pD	49.191	77,062
DIC	627.209	630.758

the *i* area is defined by

$$\vartheta_i = \exp\left(X'\beta + u_i + v_i\right). \tag{68}$$

The log log-link function is defined as

$$\log(\mu_i) = \log(E_i) + \exp(X'\boldsymbol{\beta} + u_i + v_i),$$ (69)

where $y, \boldsymbol{\beta}, E$ and ϑ are vectors of the covariate, the associated parameters, the expected number of cases, and the relative risks of TB prevalence respectively. The u_i is the county level random effect capturing the residual log RR of disease in county i. The u_i (UH) is sometime thought of as a latent variable which captures the effect of unknown or unmeasured area level covariates and v_i has a CAR model structure.

Parameter Estimation

We defined the likelihood function as

$$\ell(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}) = \prod_{i=1}^{n} \frac{(E_i \exp(\eta_i))^{y_i} \exp(-E_i \exp(\eta_i))}{y_i!} = P(y, E, \vartheta \mid \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}).$$ (70)

The prior distribution for $\boldsymbol{\beta}$ is

$$P(\boldsymbol{\beta}) = \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^P \exp\left(-\frac{1}{2}\sum_{p=0}^{P} \frac{\beta_p^2}{\tau_\beta^2}\right),$$ (71)

prior distribution for the area-specific random effect u_i is defined by

$$P(\boldsymbol{u}) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\tau_u}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{u_i^2}{2\tau_u^2}\right),$$ (72)

and prior distribution for the CAR structure v_i is

$$P(\boldsymbol{v}) = \left[v_i \mid v_{j \neq i}, \tau_v^2\right] \sim N\left(\sum_{j \neq i} \frac{w_{ij}}{w_{ij}} v_j, \frac{\tau_v^2}{w_{ij}}\right) \sim CAR(0, \tau_v^2).$$ (73)

Therefore the posterior distribution for the parameters of interest is defined as

$$P(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \tau_\beta^2, \tau_u^2, \tau_v^2 \mid y, E, \vartheta) \propto P(y, E, \vartheta \mid \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \tau_\beta^2, \tau_u^2, \tau_v^2) P(\boldsymbol{\beta}) P(\boldsymbol{u}) P(\boldsymbol{v}).$$ (74)

Therefore,

$$P(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \tau_\beta^2, \tau_u^2, \tau_v^2 \mid y, E, \vartheta) = \prod_{i=1}^{n} \frac{(E_i \exp(\eta_i))^{y_i} \exp(-E_i \exp(\eta_i))}{y_i!}$$
$$\times \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^P \exp\left(-\frac{1}{2}\sum_{p=0}^{P} \frac{\beta_p^2}{\tau_\beta^2}\right) \times \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\tau_u}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{u_i^2}{2\tau_u^2}\right) \times \left(\sum_{j \neq i} \frac{v_j w_{ij}}{w_{ij}}, \frac{\tau_v^2}{w_{ij}}\right).$$ (75)

The hyperprior disribution for the precision parameters τ_u^2, τ_v^2 and τ_β^2 are $\tau_u^2 \sim \text{Gamma}(0.5, 0.005)$, $\tau_v^2 \sim \text{Gamma}(0.5, 0.005)$ and $\tau_\beta^2 \sim \text{Gamma}(0.5, 0.01)$ respectively. The linear regression coefficient are assumed to have normal distribution defined by $\boldsymbol{\beta} \sim N(0, \tau_\beta^2)$. The τ_u^2 reflects the amount of extra-poisson variation in the data [18]. The precision variances τ_u^2 and τ_v^2 control the variability of $\boldsymbol{u}$ and $\boldsymbol{v}$ respectively. Parameter estimation was carried out using Bayesin Markov Chain Monte Carlo via Gibbs Sampling. Convergence of the MCMC was reached at 11,000 iteration after a burn-in period of 10,000 sample and thinning of every 90th element of the chain.

Posterior statistics of the CAR and the BYM model are presented in **Table 6**. **Table 6** revealed that the estimated overall relative risk effect of the CAR model is $\beta_0 = -0.177$ (95% credible interval = (−0.180, −0.174)) and BYM model $\beta_0 = -0.179$ (95% credible interval = (−0.267, −0.0908)). Each model's overall risk effect is significantly different from zero and negative. These models confirmed with the UH model that overall TB risk would be decreasing keeping all determinants of TB constant. Again, only the HIV variable is significant and positive for the CAR model and the BYM model with parameter estimates 1.812 (95% credible interval = 0.7735 - 2.758) and 1.41 (95% = 0.488 - 2.34) respectively. We therefore infer that the relative risks of TB increases as HIV prevalence increases.

As noted previously, the smaller the precision variance τ_v^2, the risk in any given area is similar to that in the neighbouring areas. The CAR model's precision variance, $\tau_v^2 = 1.56$ (95% credible interval = 0.94 - 2.19) indicates high similarities of TB relative risk between neighbouring counties than the BYM model's precision variance, $\tau_v^2 = 11.3$ (95% credible interval = 2.1 8- 40.80). High variation of TB risk exhibited by τ_v^2 in the BYM model could be due to the presence of the u_i term with precision variation $\tau_u^2 = 13.4$ (95% credible interval = 5.79 - 40.00).

Although the CAR model and BYM model each provides important information about TB relative risk behaviour, we recommend the CAR model as the best fitting spatial model to Kenya TB data since it yields lower DIC (627.21) and lower pD (49.19) than the BYM model with DIC (630.76) and pD (50.97). The BYM model, though robust, its robustness as a spatial model is lost a in situation where there is over-fitting [28]. That is, adding spatially structured extra-variability to the data when such variability doe not actually exit, conditionally on the covariates included in the model, leads to over-fitting, and may bias the estimations of the medical association between covariates and relative risk towards the hypothesis that it has no significant effect. In other words, not accounting for an actual spatial variability may lead to major biases but if spatially, variability of health indicators is completely explained by that of the socio-economic and environmental factors taken into consideration, regression residuals could results to a biased estimate of the medical association. We therefore presents detailed results of the CAR model (43). **Figure 12** shows the convergence diagnostics plots of the CAR model. The convergence diagnostics of the BYM model can be found in **Figure 13** and **Figure 14**.

Figure 15 also shows that out of the 47 counties in Kenya, 13 exhibit TB relative risk higher than the national risk (RR = 1). **Table 7** grouped counties according to their respective relative risk ranges. The pattern of risk behaviour is similar to that reported in the previous models. High TB risk occurs in the North, West, North-West and central counties of Kenya and low risk in the South-East counties except Mombasa.

Table 7 shows 4 counties (Nairobi, Mombasa, Isiolo, and Marsabit) having highest relative risk (RR > 2.0). Out of the 47 counties, 13 counties show high relative risk above 1. Counties with high TB relative risk are visually shown by the exceedence probability map **Figure 16**.

Figure 16 confirms with the UH model's results that there are 13 counties elevated risk (RR > 1). These counties exhibiting high relative risk are: Nairobi, Mombasa, Kisumu, Turkan, Migori, Homa bay, Uasin Gishu, Isiolo, Marsabit, Siaya, Tharaka-Nithi, Mandera, and Embu. Again, it can be observed that North, West, North-West and central counties of Kenya exhibit high TB prevalence and low prevalence in the South-West counties except Mombasa. **Figure 17** captures areas with potential clusters of disease risk. Clusters of TB risk are suspected in Marsabit, Embu, Migori and Kisumu.

6.4. Summary of the Spatial Models

The CAR and the BYM models are used to capture clustering or clusters information about disease risk. Each model identified HIV as a major cause of high TB prevalence in Kenya. Each model revealed significance of risk similarities between neighbouring counties. Local clusters of TB risk occurs in neighbouring counties with high TB relative risk. Though each of the CAR model and the BYM provides interesting information about Kenya's TB data, CAR model appears to provide best fit since it yields lower DIC (49.19) and lower pD (627.21) than the BYM model with DIC (50.97) and pD (630.76).

7. Discussion and Conclusion

This thesis provides a framework for application of non-spatial and spatial models for modeling and mapping TB in Kenya. We discuss the non-spatial and spatial models methodology as well as illustrate application of the non-spatial and spatial methods to TB in Kenya. We have also evaluated models performance using the DIC approach proposed by [1]. Our findings reveal that the CAR and the unstructured heterogeneity models may be suitable for modeling and mapping relative risk of TB in Kenya. **Table 8** presents comparison of the non-spatial models (PG, LN and UH), spatial model (CAR) and BYM model (combination of CAR and UH models) used in this study. Though the Poisson-Gamma (PG) model yields the lowest DIC, it does not allow for incorporation of spatial structure. The overall relative risk estimated by the PG is 0.94 (95% credible interval = 0.82 - 1.10). The Log-Normal (LN) provides specifications that can be extended to include spatial structures. The UH, CAR, and BYM models confirm that with all determinants of TB kept constant, overall relative risk of TB will be decreasing. Also, the UH, CAR, and BYM confirm HIV as a major TB determinant and that TB prevalence in
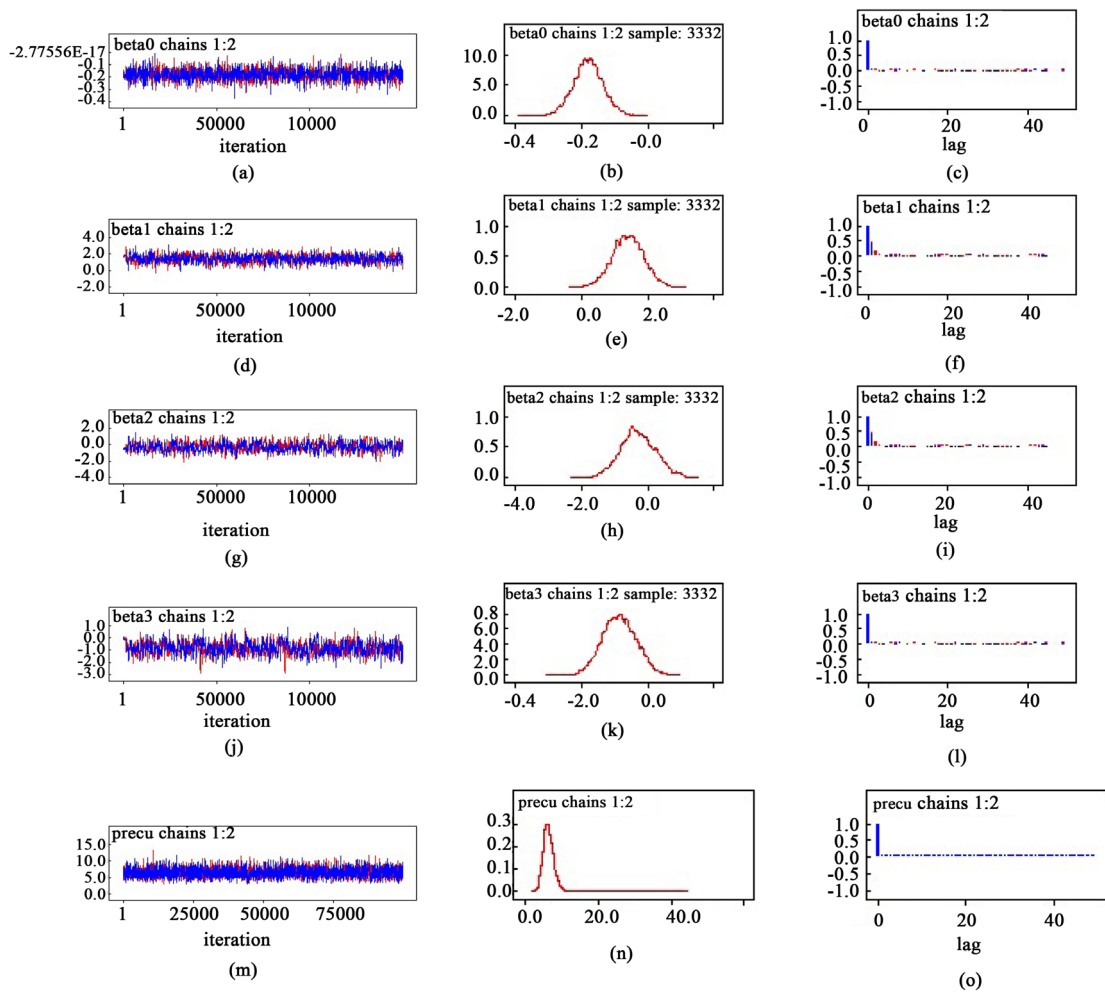
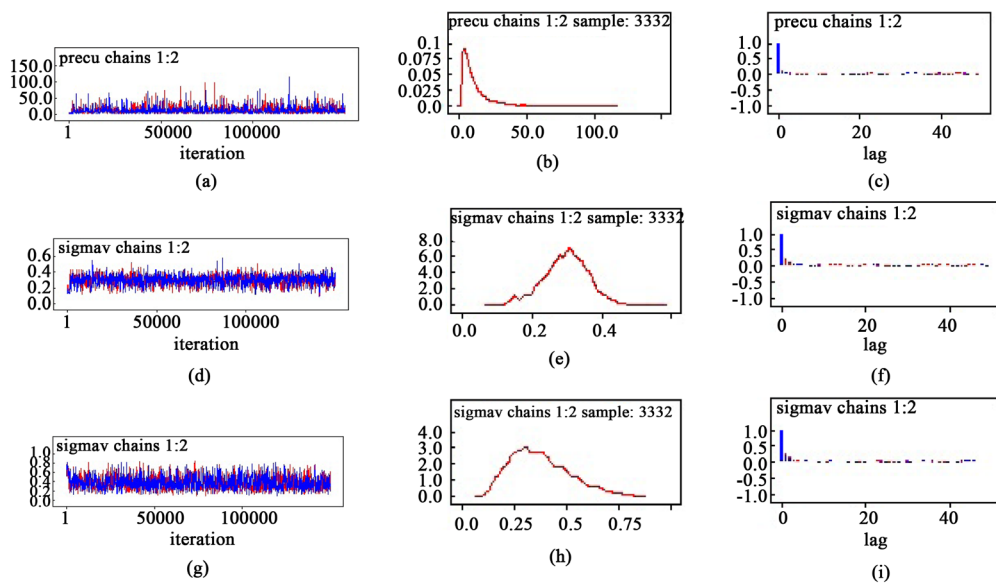**Figure 13.** BYM model: Rubin and gelman convergence diagnostics.

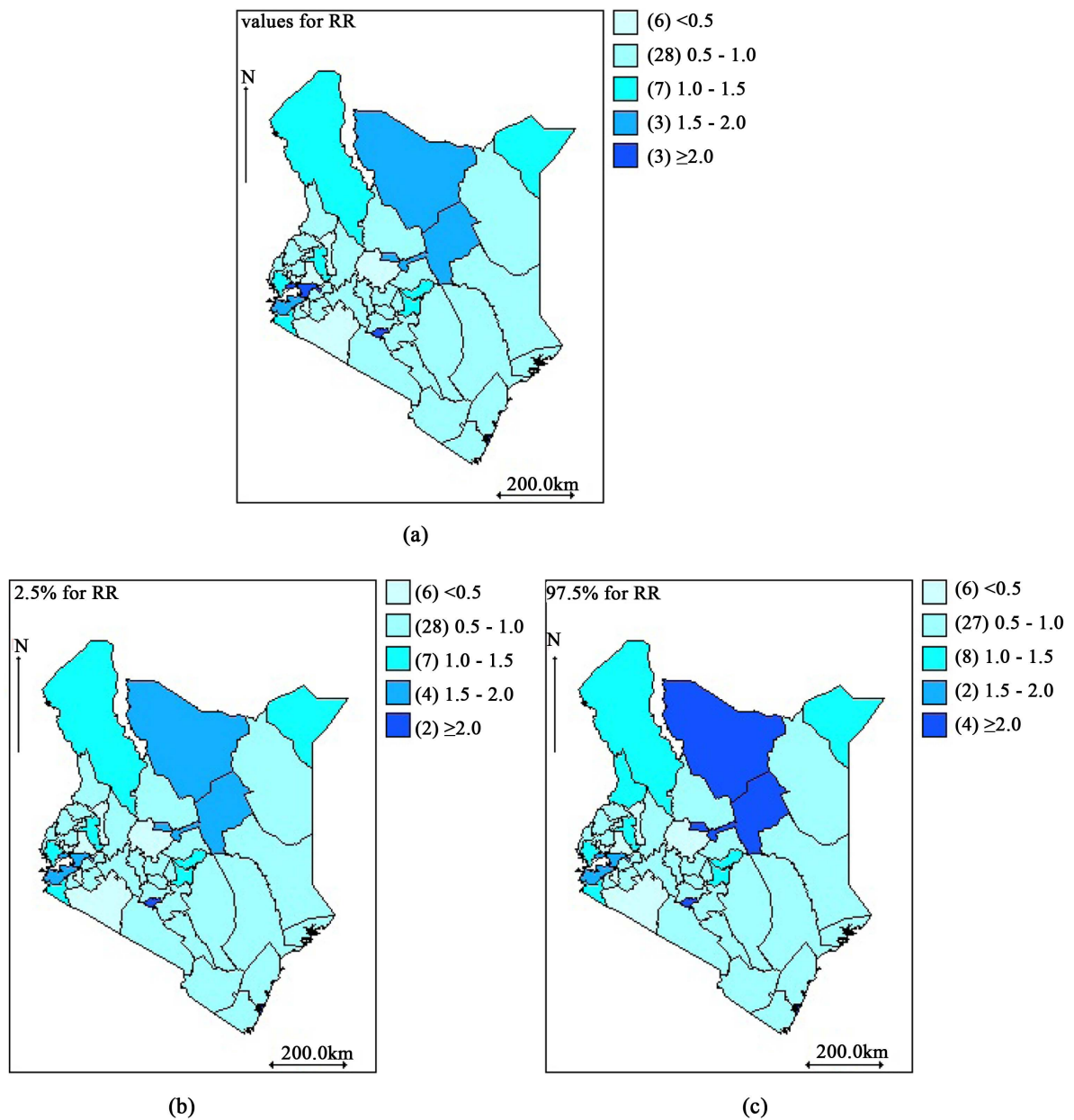**Figure 14.** BYM model: Rubin and gelman convergence diagnostics cont.

Figure 15. Kenya county level TB prevalence counts: The CAR model's posterior mean of the relative risk map (a) and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

Kenya increases with increasing HIV. Generally clustering of risk and elevated risk are observed in the North, West, North-West and the central counties of Kenya and low clustering and elevated risk in the South-West counties.

The UH model captures and displays variability of relative in the study area through the area-specific random effect whiles the CAR model and the BYM model provide evidence of risk similarities between neighboring counties. Among the LN, the UH, the CAR and the BYM models, the UH model yielded the lowest DIC (622.75), hence considered as the best fitting model when fitted to Kenya TB data for 2002-2009. However, using the acceptable criteria that a DIC difference between two models greater than 10 implied significant difference while a DIC less than 5 implied a negligible difference [29], one could use any of the non-spatial and spatial models for fitting Kenya TB data for 2002-2009 depending on the issue at hand. Although these approaches are less than ideal, we hope that our results provide a useful starting point into the development of

Table 7. Poisson log-normal model with CAR model results indicating counties with high and low TB risk.

RR > 2.0	RR: 1.5 - 2.0	RR: 1.0 - 1.5	RR < 0.5
Nairobi, 2.159 (2.145, 2.174)	Homa bay ,1.721 (1.702, 1.74)	Embu, 1.199 (1.18, 1.219)	Lakaipia, 0.458 (0.4449, 0.4714)
Mombasa, 2.383 (2.36, 2.407)	Isiolo, 1.957 (1.906, 2.01)	Mandera, 1.044 (1.02, 1.067)	Nandi, 0.4033 (0.3941, 0.4127)
Marsabit, 1.969 (1.93, 2.008)	Kisumu, 1.975 (1.955, 1.995)	Migori, 1.374 (1.357, 1.392)	Narok, 0.482 (0.4715, 0.4928)
-	-	Siaya, 1.401 (1.384, 1.419)	Nyamira (Kisii North), 0.4533 (0.4422, 0.4646)
-	-	Tharaka-Nithi, 1.064 (1.042, 1.086)	Vihiga, 0.469 (0.4581, 0.4801)
-	-	Turkan, 1.068 (1.051, 1.086)	-
-	-	Uasin Gishu, 1.182 (1.166, 1.198)	-
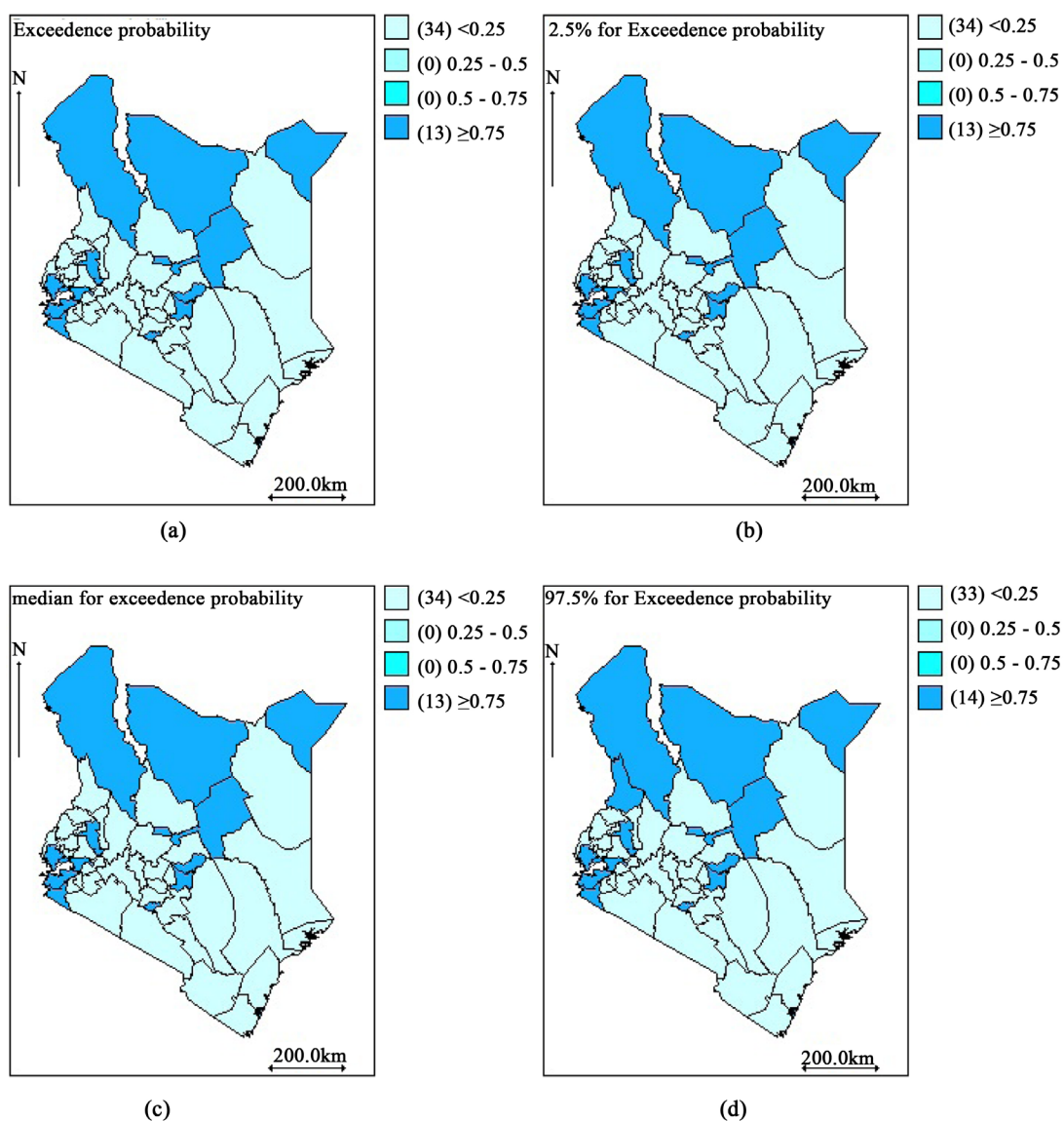
Figure 16. Kenya county level TB prevalence counts: The CAR model posterior mean of the relative risk exceedence probability map (a) and its 2.5% quantile (b), median (c) and 97.5% quantile (d).
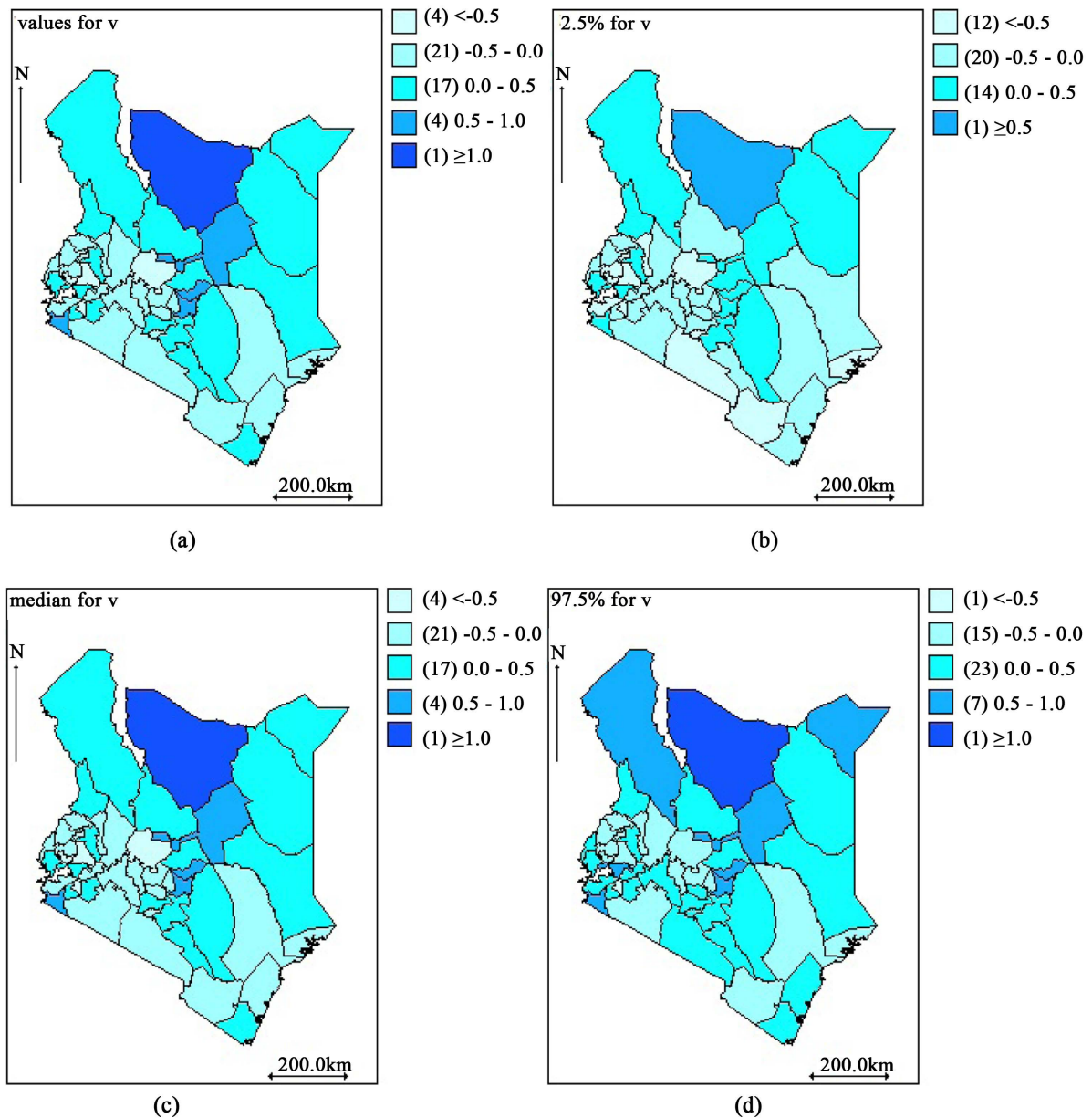
Figure 17. The correlated heterogeneity effect's posterior map (a) and its 2.5% quantile (b), median (c) and 97.5% quantile (d).

spatial methodology for modeling and mapping RR of TB in Kenya.

Modeling of risk in space and time is important but is quite a challenging task. Further research is required determined suitable spatio-temporal models for modeling and mapping relative risk of TB in Kenya. This will allow to explain evolution of the relative risk of TB in space and time.

Limitation of the study is that the specification of the adjacency matrix W with 0 and 1 in the CAR model (43) is not internally consistent in a case in which the number of neighbors varies (occurs with most irregular lattices). In the CAR model, when ρ is fixed at 1, the CAR models' specification becomes an "intrinsic" CAR model (which is prevalent in empirical studies), and requires less computation time but presents theoretical and conceptual issues that undermine its validity [30]. For instance, the precision parameter τ^2 is unknown (which is always the case), the functional from of the joint distribution of the spatial random effects (v), are not identifiable under the "intrinsic" CAR specification. Thus one cannot be confident that his/her estimates, nor

Table 8. Posterior statistics of non-spatial and spatial models.

Model indicators	PG	LN	UH	CAR	BYM
β_0	-	-	−0.177 (−0.296, −0.060)	−0.177 (−0.181, −0.174)	−0.179 (−0.267, −0.091)
a	4.72 (3.10, 6.71)	-	-	-	-
b	5.046 (3.21, 7.29)	-	-	-	-
mean	0.94 (0.82, 1.10)	−0.17 (−0.31, −0.046)	-	-	-
variance	0.20 (0.12, 0.31)	-	-	-	-
HIV	-	-	1.198 (0.493, 2.571)	1.812 (0.774, 2.758)	1.41 (0.488, 2.34)
Firewood	-	-	0.274 (−2.215, 2.144)	0.276 (−2.44, 2.822)	−0.28 (−1.29, 0.793)
five kilometer distance	-	-	−1.317 (−3.423, 1.437)	−1.505 (−4.19, 1.18)	−0.852 (−1.81, 0.124)
τ^2	-	5.012 (3.20, 7.24)	-	-	-
σ	-	0.46 (0.37, 0.56)	-	-	-
σ_v	-	-	-	0.8298 (0.675, 1.03)	0.372 (0.156, 0.678)
τ_v^2	-	-	-	1.559 (0.943, 2.194)	11.3 (2.18, 40.8)
σ_u	-	-	0.441 (0.359, 0.547)	-	0.298 (0.158, 0.416)
τ_u^2	-	-	5.324 (3.347, 7.757)	-	13.4 (5.79, 40)
pD	46.95	47.013	46.973	49.191	50.969
DIC	622.70	622.83	622.753	627.209	630.758

convergence of the parameters draw, due to potentially improper distributional assumptions. Conceptually, not including ρ in the model blurs one's estimates and can lead to counter-intuitive interpretation [30].

References

[1] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), **64**, 583-639.

[2] López-Quílez, A. and Munoz, F. (2009) Review of Spatio-Temporal Models for Disease Mapping.

[3] Currie, C.S.M., Williams, B.G., Cheng, R.C.H. and Dye, C. (2003) Tuberculosis Epidemics Driven by Hiv: Is Prevention Better than Cure? *Aids*, **17**, 2501-2508. http://dx.doi.org/10.1097/00002030-200311210-00013

[4] Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995) Bayesian Analysis of Space-Time Variation in Disease Risk. *Statistics in Medicine*, **14**, 2433-2443.

[5] Waller, L.A., Carlin, B.P., Xia, H. and Gelfand, A.E. (1997) Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association*, **92**, 607-617. http://dx.doi.org/10.1080/01621459.1997.10474012

[6] Elliott, P., Wakefield, J., Best, N., Briggs, D., *et al*. (2000) Spatial Epidemiology: Methods and Applications. Oxford University Press.

[7] Lawson, A.B. and Lawson, A.B. (2001) Statistical Methods in Spatial Epidemiology. John Wiley, Hoboken.

[8] Knorr-Held, L. (1999) Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk. *Statistics in Medicine*, **19**, 2555-2567.

[9] Lawson, A.B. (2008) Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. CRC Press..

[10] Clayton, D. and Kaldor, J. (1987) Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease

Mapping. *Biometrics*, 671-681.

[11] Besag, J., York, J. and Mollié, A. (1991) Bayesian Image Restoration, with Two Applications in Spatial Statistics. *Journal of Applied Statistics*, **5-6**, 21-59.

[12] da Roza, D.L., do Carmo Gullaci Caccia-Bava, M., Martinez, E.Z., *et al.* (2012) Spatiotemporal Patterns of Tuberculosis Incidence in Ribeirão Preto, State of São Paulo, Southeast Brazil, and Their Relationship with Social Vulnerability: A Bayesian Analysis. *Revista da Sociedade Brasileira de Medicina Tropical*, **45**, 607-615. http://dx.doi.org/10.1590/S0037-86822012000500013

[13] Randremanana, R., Richard, V., Rakotomanana, F., Sabatier, P. and Bicout, D. (2010) Bayesian Mapping of Pulmonary Tuberculosis in Antananarivo, Madagascar. *BMC Infectious Diseases*, **10**, 21. http://dx.doi.org/10.1186/1471-2334-10-21

[14] Achcar, E.Z., Martinez, A., Ruffino-Netto, J.A., Paulino, C.D. and Soares, P. (2008) A Statistical Model Investigating the Prevalence of Tuberculosis in New York City Using Counting Processes with Two Change-Points. *Epidemiology and Infection*, **136**, 1599. http://dx.doi.org/10.1017/S0950268808000526

[15] Srinivasan, R., Venkatesan, P. and Dharuman, C. (2012) Bayesian Conditional Auto Regressive Model for Mapping Tuberculosis Prevalence in India. *International Journal of Pharmaceutical Studies and Research.*

[16] Ntzoufras, I. (2011) Bayesian Modeling Using WinBUGS. Vol. 698, Wiley, Hoboken.

[17] Wakefield, N.G., Best, J.C. and Waller, L. (2000) Bayesian Approaches to Disease Mapping. *Spatial Epidemiology*: *Methods and Applications*, 104-107.

[18] Lawson, A.B., Browne, W.J. and Vidal Rodeiro, C.L. (2003) Disease Mapping with WinBUGS and MLwiN. Vol. 11. Wiley, Hoboken.

[19] Gill, J. (2002) Bayesian Methods: A Social and Behavioral Sciences Approach. Vol. 20, CRC Press.

[20] Miller, H.J. (2004) Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*, **94**, 284-289. http://dx.doi.org/10.1111/j.1467-8306.2004.09402005.x

[21] Geman, S. and Geman, D. (1984) Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

[22] Kyung, M. and Ghosh, S.K. (2009) Bayesian Inference for Directional Conditionally Autoregressive Models. *Bayesian Analysis*, **4**, 675-706.

[23] Besag, J. (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion). *Journal of the Royal Statistical Society*: *Series B*, **36**, 192-236.

[24] Cressie, N.A.C. (1993) Statistics for Spatial Data. Revised Edition, Vol. 928, Wiley, New York.

[25] Mariella, L. and Tarantino, M. (2010) Spatial Temporal Conditional Auto-Regressive Model: A New Autoregressive Matrix. *Australian Journal of Statistics*, **39**, 223.

[26] Waller, L.A. and Gotway, C.A. (2004) Applied Spatial Statistics for Public Health Data. Vol. 368, Wiley-Interscience.

[27] Banerjee, S. (2009) Spatial Autoregressive Models.

[28] Latouche, A., Guihenneuc-Jouyaux, C., Girard, C. and Hemon, D. (2007) Robustness of the Bym Model in Absence of Spatial Variation in the Residuals. *International Journal of Health Geographics*, **6**, 1.

[29] Best, N. (2011) Bayesian Hierarchical Modelling Using Winbugs.

[30] Wang, Y.Y. and Kockelman, K.M. (2013) A Poisson-Lognormal Conditional-Autoregressive Model for Multivariate Spatial Analysis of Pedestrian Crash Counts across Neighborhoods. *Accident Analysis & Prevention*, **60**, 71-84. http://dx.doi.org/10.1016/j.aap.2013.07.030