Bayesian Hierarchical Spatial and Spatio-Temporal Modeling and Mapping of Tuberculosis in Kenya

By

Abdul-Karim Iddrisu

University of KwaZulu-Natal

Submitted in partial fulfillment of a Masters by Research

in

Biostatistics

UNIVERSITY OF
KWAZULU-NATAL

INYUVESI
YAKWAZULU-NATALI

PIETERMARITZBURG CAMPUS, SOUTH AFRICA

Disclaimer

The document describes work undertaken as a masters programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Student: Abdul-Karim Iddrisu.....Date:.....

Supervisor: Professor Henry G. Mwambi.....Date:.....

Co-supervisor: Dr. Thomas N. O. Achia .....Date:.....

Acknowledgements

My first thanks goes to Allah (Subhanahu Wata-Alaa) without whose divine intervention this work would not have been successful.

Many thanks to my supervisor, Professor Henry G. Mwambi and co-supervisor, Dr. Thomas Noel Ochieng Achia for their professional guide, technical approaches, endless efforts, and urge to ensure that this work emerge one of the standard UKZN's thesis. This acknowledgement would not be complete without reference to my supervisors who made TB data available to enhance this project.

I am short of words to thank UKZN's academic and administrative members, especially that Dean and Head of School of Mathematics, Statistics and Computer Science, Professor Kesh Govinder and the academic research leader, Professor Henery G. Mwambi for their dynamic leadership and support during my research period. This is indeed a memorable turning-point of my academic career.

I dedicate this work to my lovely and caring parents, Mr and Mrs Iddrisu and the whole family for their support, advise, courage and love, to ensuring that I reach this stage of the academic ladder.

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 and space-time epidemic processes [Waller et al., 1997, Hosseini et al., 2006]. The use of frequentist methods in Biostatistics and epidemiology are common and are therefore extensively utilized in answering varied research questions. In this thesis, we proposed the Hierarchical Bayesian approach to study the spatial and the spatio-temporal pattern of tuberculosis in Kenya [Knorr-Held et al., 1998, Knorr-Held, 1999, López-Quılez and Munoz, 2009, Waller et al., 1997, Julian Besag, 1991]. Space and time interaction of risk (ψ_{ij}) is an important factor considered in this thesis. The Markov Chain Monte Carlo (MCMC) method via WinBUGS and R packages were used for simulations [Ntzoufras, 2011, Congdon, 2010, David et al., 1995, Gimenez et al., 2009, Brian, 2003], and the Deviance Information Criterion (DIC), proposed by [Spiegelhalter et al., 2002], used for models comparison and selection. 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 found clustering and heterogeneity of risk among high rate counties and the overall TB risk is slightly decreasing from 2002-2009. Interaction of TB relative risk in space and time is found to be increasing among rural counties that share boundaries with urban counties with high TB risk. This is as a result of the ability of models to borrow strength from neighbouring counties, such that near by counties have similar risk. Although the approaches are less than ideal, we hope that our formulations provide a useful stepping stone in the development of spatial and spatio-temporal methodology for the statistical analysis of risk from TB in Kenya. [1]

[1]Key words:Hierarchical Bayes, hot classes, heterogeneity, Deviance Information Criterion (DIC), Markov Chain Monte Carlo (MCMC), parsimonious, spatio-temporal, spatial, host classes,and frequentist

Contents

List of Figures viii

1 Introduction 1

1.1 Background . . . 1

1.2 Global Tuberculosis Challenge . . . 3

1.3 Literature Review of Tuberculosis . . . 6

1.4 Problem statement . . . 9

1.5 The Objectives . . . 10

1.6 Thesis Structure . . . 11

2 Preliminary Data Management and Analysis 12

2.1 Introduction . . . 12

2.2 Exploratory Analysis . . . 13

2.3 Crude Estimation of Tuberculosis Risk . . . 20

2.3.1 Estimates for Spatial Data . . . 20

2.3.2 Estimates for Spatio-Temporal Data . . . 20

3 The Bayesian Hierarchical Modelling 24

3.1 The Likelihood Function . . . 24

3.2 The Prior Distribution . . . 25

3.2.1 The Propriety . . . 25

3.2.2 Conjugate Priors . . . 25
3.2.3 Noninformative Priors . . . 26
3.2.4 Jeffery's Priors . . . 26
3.3 The Posterior Distribution . . . 28
3.3.1 The General Case . . . 28
3.3.2 The Poisson-Gamma Distribution for TB incidence Data . . . 28
3.4 Bayesian Markov Chain Monte Carlo (MCMC) Method. . . . 30
3.4.1 Markov Chain Monte Carlo Algorithm . . . 30
3.4.2 The Gibbs Sampler (GS) . . . 32
3.4.3 Assessing and Improving Markov Chain Monte Carlo Convergence . . . 33
3.4.4 Criteria for Model Selection . . . 34
3.5 Disease Cluster Detection . . . 37
3.5.1 Exceedence Probability . . . 38

4 Uncorrelated Heterogeneity Methods for Relative Risk Estimation 40

4.1 The Poisson-Gamma Model (PG) . . . 40
4.1.1 Model Description . . . 40
4.1.2 Parameter Estimation . . . 41
4.1.3 Markov Chain Monte Carlo Diagnosis . . . 42
4.1.4 Results of the Poisson-Gamma Model . . . 43
4.2 Poisson-Log-Normal Model . . . 47
4.2.1 Model Description . . . 47
4.2.2 Parameter Estimation . . . 48

4.2.3 The Likelihood Function of a Regression With Gaussian Random Effect . 48
4.2.4 Posterior Function Of Gaussian Process Regression 50
4.2.5 Markov Chain Monte Carlo Diagnosis 53
4.2.6 Results of the Poisson Log-Normal Model Without Covariate and Random Effect 54
4.2.7 Markov Chain Monte Carlo Diagnostics 57
4.2.8 Application of the Poisson-Log-Normal with UH Effect and Covariates . . 58
4.3 Summary of the Nonspatial Models 62

5 Bayesian Hierarchical Spatial Model for Use in Disease Mapping. 64

5.1 Conditional Autoregressive (CAR) Model 64
5.1.1 Parameter Estimation 68
5.2 The Besag, York and Mollié (BYM) Model 70
5.2.1 Parameter Estimation 70
5.3 Application of CAR and BYM Models to the TB Data 72
5.3.1 Markov Chain Monte Carlo Diagnostics 74
5.4 Summary of the Spatial Models 79

6 Spatio-Temporal Modelling 81

6.1 Methodology 82
6.2 Bernardinelli et al. (1995) Approach 83
6.2.1 Parameter Estimation 84
6.3 Waller et al. (1997) approach 84
6.3.1 Parameter Estimation 85

6.4 Knorr-Held and Rasser, 2000 86

6.4.1 Parameter Estimation 87

6.5 Application of Spatio-Temporal Models 90

6.5.1 Type IV interaction Model 92

6.6 Summary of Spatio-Temporal Models 95

7 Discussion and Conclusion **96**

References **118**

List of Figures

2.1 Box Plots of TB Determinants of Tuberculosis in Kenya 16

2.2 Box Plots of TB Determinants of Tuberculosis in Kenya 17

2.3 Box Plots of Observed Tuberculosis Cases and Population Size of Kenya From 2002-2009 18

2.4 Box plot of County Level TB Standardized Mortality Ratio for 2002-2009 21

4.1 Poisson-Gamma Model: Convergence Diagnosis of Markov Chain Monte Carlo. . 42

4.2 Kenya county level Standardized Mortality Ratio's maps: (4.2a) The mean of the SMR and its 2.5% quantile (4.2b), median (4.2c) and 97.5% quantile (4.2d). . . 44

4.3 Kenya county level Poisson-Gamma posterior mean relative risk maps: (4.3a) The mean of the posterior relative risk and its 2.5% quantile (4.3b), median (4.3c) and 97.5% quantile (4.3d). 45

4.4 Poisson-Gamma posterior relative risk exceedence probability map: Row-wise from the top left Figure: (4.4a) the posterior mean relative risk exceedence probability and its 4.4b 2.5% quantile for relative risk, (4.4c) median for the relative risk and (4.4d) 97.5% quantile for the relative risk. 46

4.5 Poisson Log-Normal Model: Convergence Diagnosis of Markov Chain Monte Carlo 53

4.6 Kenya County Level TB Prevalence Counts: Poisson Log-Normal model Without covariate and random effect. (4.6a) The posterior relative risk map and its 2.5% quantile (4.6b), median (4.6c) and 97.5% quantile (4.6d). 55

4.7 Kenya County Level TB Prevalence Counts: Poisson Log-Normal model Without covariate and random effect. (4.7a) The posterior relative risk exceedence probability map and its 2.5% quantile (4.7b), median (4.7c) and 97.5% quantile (4.7d). 56

4.8 Poisson Log-Normal Model with Uncorrelated Heterogeneity (UH) Effect: Convergence Diagnosis of Markov Chain Monte Carlo 57

4.9 Kenya County Level TB Prevalence Counts: UH smooth relative risk map (4.9a) and its 2.5% quantile (4.9b), median (4.9c) and 97.5% quantile (4.9d). 59

4.10 Kenya County Level TB Prevalence Counts: UH smooth relative risk exceedence probability map (4.10a) and its 2.5% quantile (4.10b), median (4.10c) and 97.5% quantile (4.10d). 61

4.11 Area-Specific Random Effect: The posterior map (4.11a) and its 2.5% quantile (4.11b), median (4.11c) and 97.5% quantile (4.11d). 62

5.1 Poisson Log-Normal Model with CAR Model: Convergence Diagnosis of Markov Chain Monte Carlo 74

5.2 Kenya county level TB prevalence counts: The CAR model's posterior mean of the relative risk map (5.2a) and its 2.5% quantile (5.2b), median (5.2c) and 97.5% quantile (5.2d). 76

5.3 Kenya county level TB prevalence counts: The CAR model posterior mean of the relative risk exceedence probability map (5.3a) and its 2.5% quantile (5.3b), median (5.3c) and 97.5% quantile (5.3d). 78

5.4 The Correlated Heterogeneity effect's posterior map (5.4a) and its 2.5% quantile (5.4b), median (5.4c) and 97.5% quantile (5.4d). 79

6.1 Type IV interaction posterior mean of the relative risk maps for 2002-2009. . . . 92

6.2 Type IV interaction temporal trend Posterior mean of the relative risk 2002-2009 93

6.3 Type IV interaction trend of TB prevalence rate in Kenya from 2002-2009 94

6.4 Type IV temporal trend and area posterior mean relative risk. 94

7.1 BYM Model: Rubin and Gelman Convergence Diagnostics 100

7.2 BYM model: Rubin and Gelman Convergence Diagnostics 101

7.3 Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009 102

7.4 Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009 continue 103

7.5 Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009 continue 104

7.6 Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009 continue 105

7.7 Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009106

7.8 Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009 continue 107

7.9 Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009 continue 108

7.10 Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009 continue 109

List of Tables

2.1 Summary Statistics for variables in the data set 13

2.2 County Level Population Size Summaries for 2002-2009 14

2.3 Summarises Statistics of County Level TB Cases From 2002-2009 15

2.4 Counties With Extreme TB cases From 2002-2009 19

2.5 Counties With Extreme Population For 2002-2009 19

2.6 Summary of County Level TB Standardized Mortality Ratio for 2002-2009 . . . 22

2.7 Summary of Period Specific TB Standardized Mortality Ratio 22

4.1 Posterior Statistics of the Poisson-Gamma Model. 43

4.2 Posterior Statistics of the Poisson Log-Normal Model. 54

4.3 Posterior Statistics of the Poisson Log-Normal Model with UH Effect. 58

4.4 UH results indicating counties with high and low TB risk. 60

5.1 Posterior Statistics of the CAR and BYM Models 72

5.2 Poisson Log-Normal model with CAR Model results indicating counties with high and low TB risk. 77

6.1 Spatio-Temporal models Deviance Summaries 90

6.2 Interaction Types Deviance Summaries 91

6.3 Interaction Type IV Posterior Statistics 91

7.1 Posterior Statistics of Non-spatial and Spatial Models 96

Chapter 1

Introduction

1.1 Background

Tuberculosis (TB) is a contagious disease caused by a germ known as "Mycobacterium tuberculosis" ("M. tuberculosis"). TB affects virtually every part of the body, particularly the lungs. The infection in the lungs is referred to as pulmonary infection and is the major cause of TB transmission from one person to another. Just like the influenza, M. tuberculosis is transmitted when Active TB person (TB infected person) exhales the droplets of nuclei carrying the tubercle bacilli, and a susceptible individual inhales this droplet from the air. This droplet eventually reaches the lungs and spread throughout the body. The body's natural defence mechanism limits its multiplication and some remain dormant ("sleep") but viable, creating a condition refer to as Latent TB infection (LTBI) [CDC Training, 2000]. A person in LTBI condition has a low chance becoming infected with TB (active TB infection (ATBI)). The most vulnerable individuals are children under four years, HIV/AIDS patients, cancer and diabetes patients [CDC Training, 2000].

General symptoms of tuberculosis are: fever, chills, night sweats, loss of appetite, weight loss and fatigue and significant finger clubbing. Pulmonary symptoms include chest pains and prolonged cough producing sputum. In some instances, TB patients may cough up blood.

Generally, TB is diagnosed with microscopic examination and chest radiography and tubercle skin test purposely for LTBI diagnosis [CDC Training, 2000, Wilce et al., 2004]. Active TB infection is most often treated with a combination four drugs for at least six months [CDC Training, 2000]. Treatment for a person infected with both TB and HIV (co-infection) and Multiple Drug Resistance TB (MDR-TB) is very complicated and expensive [CDC Training, 2000, Wilce et al., 2004]. It is recommended that those with high risk of developing active TB condition

such HIV/AIDS patients and those in contact with TB patients should be tested and treated accordingly.

Tuberculosis is a global public health threat and concern. Kenya with a population of 40 million is one of the countries with the highest TB burden. Despite several measures by the Government of Kenya, various TB research institutions, and National TB programmes for combating TB in Kenya, Kenya is ranked 13^{th} among the 22 TB high-burden countries. The pandemic in Kenya constitutes 80% of the global TB incidence [Figueroa-Munoz et al., 2005]. A study conducted in Kenya's Homa Bay District hospital estimated 17% (1,122) of those who visited the TB lab between August 2005 and 2007 of which females and males constituted 51.3% and 48.7% respectively had TB [Emmanuel and Shitandi, 2010]. The increasing Beijing/W M.tuberculosis genotype was identified in Kenya (Nairobi), which is highly likely to increase TB cases if it is highly transmissible [Glynn et al., 2006].

Sub-Saharan African countries have shown noticeable decrease in HIV prevalence rate from 10% in the late 1990s to 6.1% in 2005 [Sánchez et al., 2009]. However, more research is require for HIV prevention to control TB since HIV and TB are epidemiologically associated [De Cock and KM, 1999]. Observed co-dynamics suggest that the two diseases are directly related at the population level [Lawn and Wilkinson, 2006] and also within the host [Ramkissoon, 2012].

Several factors account for TB persistence in Kenya, some of which include: cultural beliefs [Ward et al., 1997], inadequate TB health care facilities [Ayisi et al., 2011], poverty/financial constraint and socio-economic factors [Schwarz, 1980], unfamiliarity with TB causes and symptoms Ayisi et al. [2011], hereditary predisposition, alcohol, smoking, witchcraft [Baliddawa et al., 2003], HIV/AIDS [Ferreira Gonçalves et al., 2009], charcoal burning and living with domestic animals, use of common feeding utensils [Ayisi et al., 2011], and self-administration of MDR-TB medicine [Sinha and Tiwari, 2010]. HIV/AIDS is the main cause of TB increase in Kenya, and its epidemics had resulted to 10-fold increase in TB cases [Marum et al., 2006]. The promising and assuring fact is that TB is curable.

Great strides have been made in the fight against TB in Kenya. Some of the initiatives include: the establishment of the Kenyan Medical Research Institute and Center for Disease Control and Prevention (KEMRI/CDC) which supports Health and Demographic Surveillance System (HDSS)

in Kenya for TB control [Hawley et al., 2003]. The Ministry of Health and APHIA II Operational Research (OR) project which have integrated TB screening and referral services into postnatal care in five health facilities in Kenya [Marseille et al., 2011]. Upon TB occurrence in Kenya, the Kenya Tuberculosis Investigation Center (KTIC) was formed in the mid fifties. They work in collaboration with the British Medical Research Council (BMRC) and WHO for TB survey and clinical diagnosis [Schwarz, 1980]. Other bodies responsible for TB control in Kenya are: the Kenya association for the Prevention of Tuberculosis (KAPT) [Schwarz, 1980], the Tuberculosis and Respiratory Infection Unit (TRIU) [Schwarzer et al., 2002], and the President's Emergency Plan For AIDS Relief (PEPFAR) and the Global Health Initiative (GHI) in 2009 by His Excellency President Obama's administration [Fleischman, 2011].

1.2 Global Tuberculosis Challenge

According to WHO, despite the fact that TB is curable, more than 5,000 people died every day, (that is 2-3 million per year) [Wilce et al., 2003]. Global TB incidence is estimated at 1.6 million each year, with one third of the world's population estimated to have TB infection [Wilce et al., 2003]. TB became global concern following the emergence of MDR-TB in many parts of the world including: Russia, Latvia, Estonia, Argentina, the Dominican Republic, and Ivory Coast, with about 50 million people infected with MDR-TB worldwide [Wilce et al., 2003]. Globally, WHO reported that TB is the leading Cause Of Death (COD) in people living with HIV/AIDS. One in four deaths among people living with HIV/AIDS is as a result of TB, especially, poor socio-economic areas.

Global views concerning TB cause and symptoms vary among countries and ethnic groups. Philippines TB patients attribute TB symptoms to drinking and smoking [Auer et al., 2000], Igbo of Nigeria TB patients attribute TB cause to eating beef and other high protein foods [Enwereji, 1999], Bostwana patients associate TB symptoms to hard work in mines, drinking and smoking [Mazonde and Mazonde, 1999], Vietnamese refugees in the US said TB is caused by hard manual labour, smoking, alcohol, poor nutrition, and germs [Carey et al., 1997], Malawi's TB patients viewed TB causes to as a result of adultery, germs, alcohol abuse, "wrong" food, stagnant water,

and witchcraft [A et al., 2000], TB and all other disease in general was viewed to be caused by imbalances in behaviour or diet in Ethopia [Vecchiato, 2008], and Xhosa-speaking in South Africa attributed TB causes to lack of hygiene and witchcraft [Wilkinson et al., 1999]. The people in rural Haiti believed that TB is caused by sorcery [Farmer et al., 1991]. Numerous factors are common to those listed above that may hinder TB control [Mazonde and Mazonde, 1999].

Stigmatization accelerates TB's spread and increase control difficulties. TB patients tend to hide their status from their families [Wilce et al., 2003]. Stigma to TB was also confirmed in Vietnamese [Carey et al., 1997], Mexico immigrants in California [Rubel and Garro, 1992], in India [Uplekar and Rangan, 1996], in South Africa Zulus [Rubel and Garro, 1992]. There seem to be acommon global view about the causes and symptoms of TB most of which causes their delay in seeking health care.

Various models and TB antibiotics have been developed by almost all affected countries in connection with the WHO Stop TB strategy [Raviglione, 2007] and Stop TB partnership's Global Plan to Stop TB [Raviglione, 2007]. In 2008, 19% (180) countries were reporting and all the 22 TB high-burden countries are implementing the DOTS component of the WHO Stop TB strategy [K et al., 1999]. The implementation of standard TB diagnostic and treatment approaches by WHO had cured 36 million people between 1995-2008, preventing up to 6 million deaths [Lönnroth et al., 2010]. Regarding the objective of the millennium goal, TB control would have been achieved in 2004, and even the more important long-term elimination of TB set for 2050 is unlikely to be achieved with the current strategies [Lönnroth et al., 2010].

Globally, 46% TB patients are receiving HIV anti-viral drug and 77% have started the co-trimozazole preventive treatment in 2010 [Organization et al., 2011]. New TB diagnostic tool (Xpert MTB/RIF) which test for TB in 100 seconds is currently use by 26 countries since July 2011. This tool was endorsed by WHO and 145 countries are allow to purchase the kits at affordable price [Organization et al., 2011]. Internal funds for TB control among affected countries have risen to an estimated value of 85%, but most low income countries still depend on external donors. External donors are provided 82% International TB funding for 2012.

Social researchers emphasized on the need to go beyond biomedical model for TB control [Wilce et al., 2003]. More research is required on DOTS programme [Wilce et al., 2003]. TB is a

global public health issue and if no tangible measures are taken, TB is likely to escalate [Porter et al., 1999]. There are "point-of-case" test under way, 10 TB drugs on trial, and 10 vaccines candidates for prevention of TB in phase I or phase II trials [Organization et al., 2011].

Considering the complex nature of TB, theories regarding its origin and global transition to the current state continue to change in correspondence with new archaeological discoveries and evolution of molecular technology [Davis and AL, 2000]. With diverse and extensive TB research and studies, scientists have hypothesised that M.tuberculosis is associated with mycobacterium, M. bovis, coincident with domestication of cattle by humans at approximately 15,000 years ago [Daniel and TM, 2000]. TB was identified in Egyptian mummies dating 5,400 years ago [Thomas, 1997].

Tuberculosis was associated with romanticism during the 19th century [Ott, 1996]. The search to understand the pathophysiological and clinical manifestation of TB begun during the 17th, 18th, and 19th centuries [Reichman and Hershfield, 2000]. Robert Koch identified the tubercle bacilli in 1882 as the main cause of TB and established it as an infectious disease [Reichman and Hershfield, 2000]. In the 1940s, scientist discovered that the antibiotic, streptomycin, killed M. TB; however, the bacilli has the ability to develop resistance when only streptomycin was used. By 1950, combined drug treatment for TB was introduced [Davis and AL, 2000].

My understanding is that treatment of TB is not a problem. Probably the problem is adherence to TB treatment and the impact of HIV.

1.3 Literature Review of Tuberculosis

There exist a vast amount of literature concerning the development and application of disease mapping approaches some of which can be found in [Bernardinelli et al., 1995, Waller et al., 1997, Wakefield et al., 2000, Lawson and Lawson, 2001, Knorr-Held, 1999]. The global spread of infectious diseases threatens the well-being of not just the human population but 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 subject to complication brought about by heterogeneity that could be of spatial or non-spatial nature [López-Quılez and Munoz, 2009, Currie et al., 2003]. Ignoring these complexities is likely to lead to incorrect inference and erroneous conclusion [López-Quılez and Munoz, 2009, Currie et al., 2003]. Disease data can be case-event data at locations affected individuals are found or counts from non-overlapping regions. Since the data provided for this study only had county level spatial resolution, the approach would be to analyse the data as regional or county specific count data.

Spatial and spatio-temporal distribution of a disease is often understood through application of statistical methods to the data and creating maps that visually describes spatial and spatiotemporal variation of disease risk [Currie et al., 2003]. However, disease counts maps are subjected to numerous problems. One such problem is the Modifable 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 has variety of special cases one of which is ecological or medical bias where the problem is whether inference can be made at the individual level from aggregate data. The question often asked is, can we make inference from county or region level analysis to the individual level. 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. Multiscale Analysis can also be used to addressed MAUP. This analysis concerns spatial units that are completely matched when aggregated [Lawson, 2008].

Also differences in population between regions results to differences in variance of regional estimates. This problem is addressed by employing a hierarchical Bayesian model that smooths the risk from neighbouring regions and clearly accounts for population difference by using a Poisson distribution for outcomes.

Bayesian methods are widely used in disease mapping. Clayton and Kaldor [1987] 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 hyperparameters.

Julian Besag [1991] is the first example of fully Bayesian disease mapping. They used the convolution prior model described in Section 5.1 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 Julian Besag [1991], the fully Bayesian model produced more accurate estimates than the EB produce of Clayton and Kaldor.

Modelling of count data in space-time has received considerable development. Bernardinelli et al. [1995] is the first example of modelling count data in space-time. They assumed Poisson likelihood model for the count data where the log of the relative is the focus of modelling.

Waller et al. [1997] proposed a spatio-temporal model that have several similar features as the BYM or convolution model. This model was based on convolution prior and allowed for each period to have separate spatial and nonspatial effect. They assumed that the covariates effects are constant over the study period and that disease counts followed a Poisson distribution. They fitted this model lung cancer deaths in 88 Ohio counties for the year 1968-1988. Each year was treated as a separate time period. Some of their remarkable findings were an overall trend of increasing lung cancer deaths and also increase in both spatial clustering and uncorrelated heterogeneity. They related these findings to "increasing evidence of clustering among the high rate counties, but with higher rates increasing and lower rates constant" That is increasing heterogeneity over the study period [Norton, 2008].

Knorr-Held et al. [1998] proposed a spatio-temporal model which included both the convolution prior on space and also a similar prior for temporal trends. They extended the Waller et al. model by assuming that the spatial terms were constant over the study periods. This followed after they stated that "Waller et al. formulation in principle does allow exploration of additionally varying risk factors within each year, but is built on the premise that temporal smoothing is unnecessary, treating time as essentially exchangeable". This model was applied to the same Ohio lung cancer data, but it was not clear that it revealed additional features of the data [Norton, 2008].

Roza et al. [2012] 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, South-east Brazil between 2006-2009 and to evaluate TB risk determinants. The study revealed that TB rates are correlated with measures of income, education and social vulnerability. They stated 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.

Randremanana et al. [2010] 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.

Achcar et al. [2008] 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 programmes 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.

Srinivasan and Dharuman [2012] 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.

1.4 Problem statement

One-third of the world's population is infected with *M. tuberculosis*, leading to 3 million deaths each year [Murphy et al., 2003]. More investigation on TB dynamics need to be explored to establish spatial, spatio-temporal distributions and cause of TB pandemics. Disease risk often exist in space and time and would therefore be properly understood if studied in space and time.

Statistical models are one of the tools that have been used to successfully analyse and understand the possible trends exhibited in infectious disease. Modelling of disease risk in space and time is quite challenging [López-Quılez and Munoz, 2009]. This type of analysis is often posed with problems since the number of disease cases and their associated population at risk in any single unit of space and time are too small to produce a reliable estimate of the underlying disease risk without "borrowing information" from neighbouring regions.[Knorr-Held et al., 1998].

1.5 The Objectives

This project is aimed at modelling and mapping TB in Kenya and to propose or suggest models for modelling and mapping TB in Kenya.

The specific objectives of this study are:

1. To review statistical methods to handle spatial and spatio-temporal models used in disease mapping
2. To identify appropriate Spatial models for modelling and mapping TB risk in Kenya over the period 2002-2009 using routine surveillance data from Kenya DHS.
3. To identify appropriate spatial-temporal models for modelling and mapping TB risk in Kenya over the period 2002-2009 using routine surveillance data from Kenya DHS.

1.6 Thesis Structure

Chapter 2 presents a summary of the data set used in this study and explains the standardization methodology used to compute the expected number of disease cases and standardized mortality ratio (SMR).

In Chapter 3, we give Bayesian methodology and techniques based on which parameters are obtained.

In Chapter 4, we present nonspatial models used in disease modelling and mapping and their applications to Kenya TB prevalence data for 2002-2009.

In Chapter 5, we discuss the Bayesian hierarchical spatial model used in disease modelling and mapping and their applications to Kenya TB prevalence data for 2002-2009.

In Chapter 6, we discuss some of the spatio-temporal models used in disease modelling and mapping and their application to Kenya TB prevalence data for 2002-2009. Finally, in Chapter 7, we give the discussion and conclusions of the thesis.

Chapter 2

Preliminary Data Management and Analysis

2.1 Introduction

Kenya is located in the Eastern part of Africa and is divided into 8 provinces and 47 administrative counties. Kenya share borders with Tanzania at the south, Uganda at the West, Ethopia at the North, Somalia at the North-East and Southern Sudan at the North-West. According to Kenya National Bureau of Statistics (KNBS), Kenya's population was estimated at approximately 39.5 million in 2011. The Gross Domestic Product (GDP) in Kenya was worth 35.557 billion US dollars in 2011/2012.

According to Daima Kenya statistics, HIV Prevalence rate in Kenya was estimated at 6.3% in 2012. Approximately 1.5 million people in Kenya are living with HIV. More then 70% of those infected with the HIV virus live in the rural areas whiles only 30% live in the urban areas. The number of people living with HIV in Kenya has dropped from 13% in 2000 to 6.3% in 2012.

The data used in this study is routine data from Kenya DHS. It contains 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 summarises.

Some of the determinants of TB that were recorded are:

1. HIV prevalence
2. Poverty prevalence

3. Illiteracy
4. Population less then 5km to health facility
5. Firewood
6. Altitude and
7. Mean house hold size

Summaries of all variables that constitute the data are presented in the next section.

2.2 Exploratory Analysis

Table 2.1: Summary Statistics for variables in the data set

Variables	No. of counties	Mean	SD	Median	Min	Max	95% CI
TB Cases	47	17830	22348.07	12531	1348	149600	(7200,20560)
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)
House hold 5km 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)

Table 2.1 presents summaries of some determinants of TB in Kenya. The mean and the median estimates of TB cases and illiteracy revealed that most of their values are concentrated at the high scale. Following the same analogy, the HIV prevalence, Proportion of poor people in the population and the mean house hold size variables have almost equal values at the low and high scale. These variables are almost symmetric or normally distributed. Finally, the variables, percentage of people who are at 5km distance away from hospital and the percentage of those who use firewood have most of their data values concentrated at the low scale. Variable distribution

and presences of outliers are visually shown with the help of box plot shown in Figure 2.1 and Figure 2.2.

Table 2.2: County Level Population Size Summaries for 2002-2009

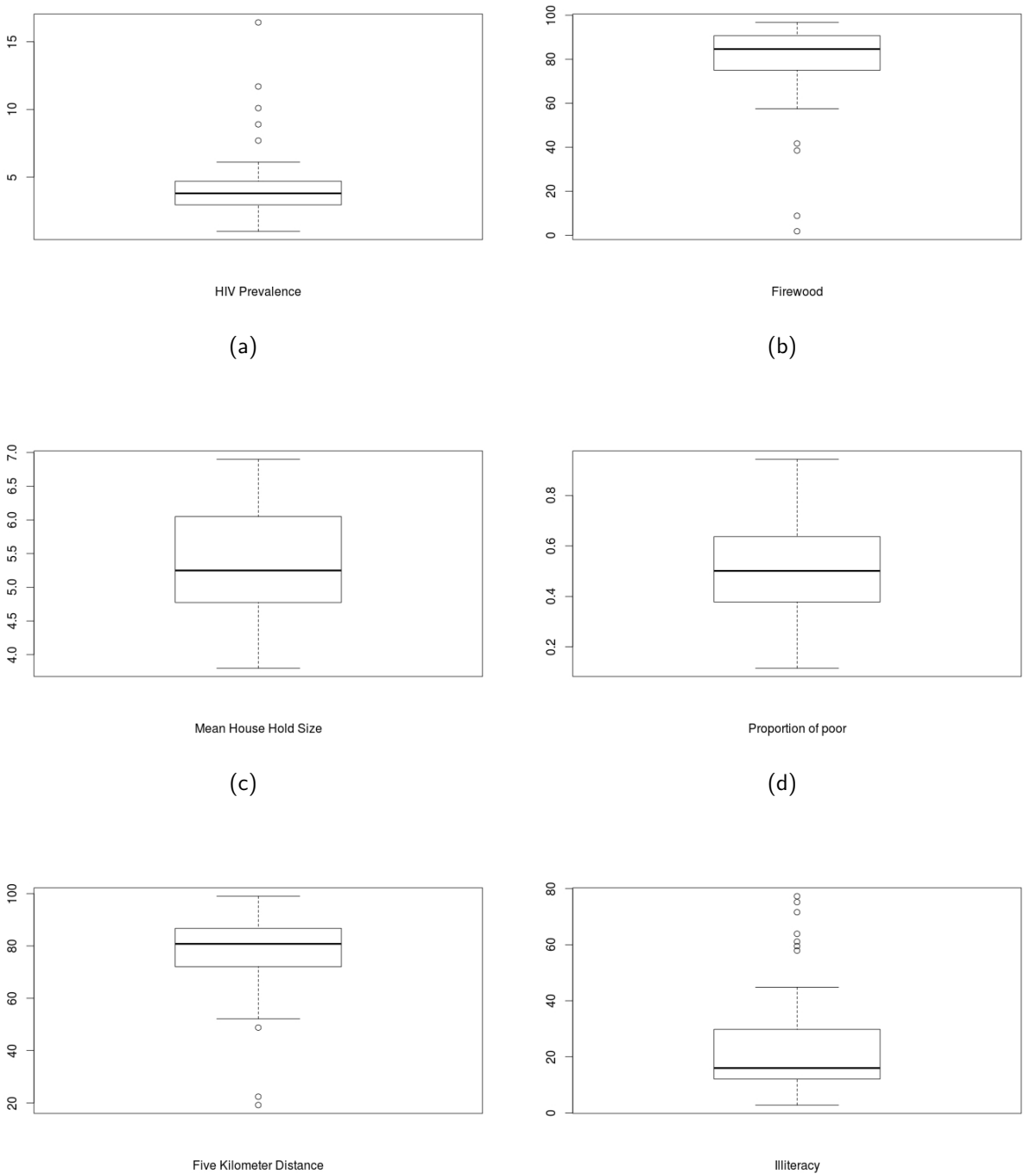
Year	N	Mean	SD	Median	Min	Max	95% CI
2002	47	685,586	435,819.5	604,298	83,985	3,034,397	(407424,869752)
2003	47	705,776	450,910.0	625,506	86,443	2,600,859	(420676,890554)
2004	47	727,220	466,775.6	647,886	89,058	2,710,706	(434932,912918)
2005	47	747,631	481,824	669,403	91,549	2,815,838	(449244,934009)
2006	47	768,909	497,822	691,637	94,150	2,924,309	(462472,956223)
2007	47	791,147	514,520	714,590	96,851	3,034,397	(475354,979870)
2008	47	814,422	531,548	738,321	99,662	3,146,303	(490594,1005057)
2009	47	838,793	548,905	762,870	102,593	3,260,124	(506907,1031868)

Table 2.2 presents summaries of Kenya's population size for 2002-2009. The mean county level population size stood out to be $685,586\,(407424,869752)$ in 2002 increasing to $838,793\,(506907,1031868)$ in 2009. The mean and the median showed that most of the data values for 2002-2009 are concentrated at the high scale. Population distributions and presence of possible outliers are visually shown with the help of box plot shown in Figure 2.3

Table 2.3: Summarises Statistics of County Level TB Cases From 2002-2009

Year	N	Mean	SD	Median	Min	Max	95% CI
2002	47	1,747	2,409.296	1,053	111.0	15,979.0	(679.5,2006.5)
2003	47	2,028	2,761.937	1,355	154.0	18,360.0	(746.5,2386.0)
2004	47	2,249	2,954.713	1,497	138.0	19,871.0	(928.5,2620.0)
2005	47	2,302	2,913.039	1,549	124.0	19,487.0	(938.5,2737.5)
2006	47	2,452	2,934.737	1,757	172	19,472	(1044,2844)
2007	47	2,416	2,836.123	1,988	177.0	18,901.0	(966.5,2786.5)
2008	47	2,291	2,790.387	1,676	223.0	18,589.0	(850.5,2623.0)
2009	47	2,346	2,843.675	1,700	249	18,984	(896,2724)

Table 2.3 presents summaries of TB cases in Kenya for 2002-2009. The mean county level TB cases is estimated at $1,747\,(679.5, 2006.5)$ in 2002 and increased to 2006 at an estimated value of $2,452\,(1044, 2844)$. TB cases decreased from 2007-2008 with corresponding values of $2,416\,(966.5, 2786.5)$ and $2,291(850.5, 2623.0)$ respectively. TB cases slightly increased in 2009, estimated at $2,346\,(896, 2,724)$. The data distribution and the presence of outliers in the data are diagnosed with the help of box plots shown in Figure 2.3.

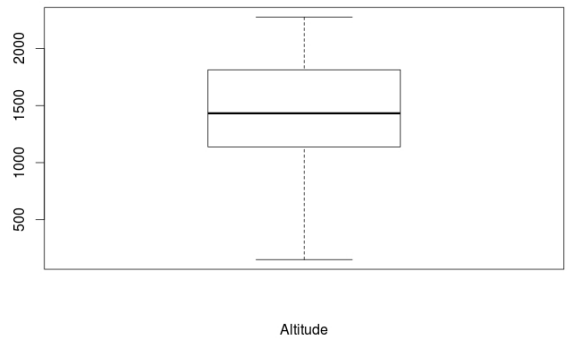


(a) (b)

(c) (d)

(e) (f)

Figure 2.1: Box Plots of TB Determinants of Tuberculosis in Kenya



(a)

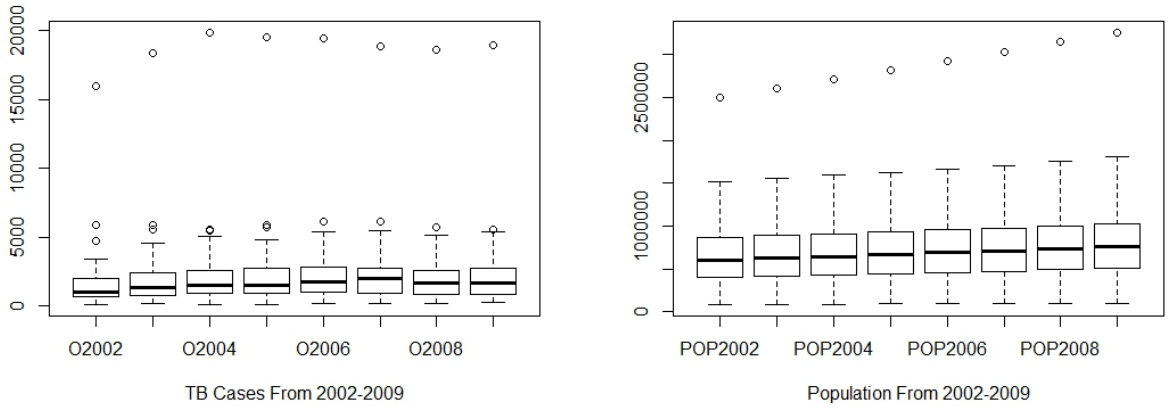
Figure 2.2: Box Plots of TB Determinants of Tuberculosis in Kenya

Figure 2.1 shows box plots of TB determinants. The box plot is one of the most widely used statistical techniques to identify patterns that may be hidden in a group of numbers or data set. The box plot uses the median (horizontal middle thick line), the quartiles (ends of the box) and the lowest and highest data points (shown by "tail" extended from above and below the box). The box plot is widely used to identify extreme low or high values in a data set. These extreme values are referred to as outliers. Visually, illiteracy variable has most of its values concentrated at the high scale. Hence, illiteracy data are positively skewed. The HIV prevalence, Altitude, Proportion of poor people in the population and the mean house hold size variables have almost equal values concentrated at the low and the high scale. These variables are symmetrically or normally distributed. Also, the variables, percentage of people who are at 5km distance away from hospital and the percentage of those who use firewood, have most of their data values concentrated at the low scale and are said to be negatively skewed.

Extreme points shown in Figure 2.1 are called outliers. Five high extreme outliers are observed in the HIV prevalence variable. These outlier correspond to data values from Mombasa (11.7%), Nairobi (10.1%), HomaBay (16.43%), Kissumu (8.9%) and Siaya (7.7%). Four low extreme values are observed in the percentage of those who use firewood variable. These low extreme values correspond to Mombasa (8.8%), Nairobi (1.8%), Nakuru (41.7%) and Kajiado (38.5%). Seven High extreme values are also observed in the percentage of illiteracy variable. These extreme values correspond to Mandera (71.6%), Garissa (57.9%), Marsabit (63.9%), Samburu

(61.1%), Tana River (59.5%), Turkana (77.3%) and Wajir (75.2%). Finally, four low extreme values are observed in the percentage of people who are 5km distance away from hospital variable. These low extreme values correspond to Isiolo (48.8%), Mombasa (22.4%), Nairobi (19.2%) and Taita Taveta (52.2%) and no outliers are observed in the mean house hold size, Altitude and proportion of population who are poor.

Among these variables, method of variable selection was carried out on them to identify and select variables that are significant for further analysis. Only HIV prevalence, percentage of people who are five km distance away from health facility and percentage of people who use firewood were identified as significant. These variables are considered for further analysis in Section 4.2 and chapter 5



(a) (b)

Figure 2.3: Box Plots of Observed Tuberculosis Cases and Population Size of Kenya From 2002-2009

Figure 2.3a visually presents distribution and presence of outliers in TB observed cases. Table 2.4 below shows counties that correspond to the high extreme values of TB cases for 2002-2009.

Table 2.4: Counties With Extreme TB cases From 2002-2009

Year	Nairobi	Mombasa	Kissumu
2002	15,979	5,889	4,753
2003	18,360	5,919	5, 581
2004	19,871	5,549	5,446
2005	19,487	5,869	5,722
2006	19,472	6,130	6,177
2007	18,901	6,157	-
2008	18,589	5,711	-
2009	18,984	5,554	-

Figure 2.3a confirmed an observed increase in TB cases from 2002-2006, decrease from 2007-2008 and a slight increase in 2009. TB cases for 2002-2009 have most of their values concentrated at the high scale, hence, are said to be positively skewed.

Figure 2.3b shows that most of the population values for 2002-2009 are concentrated at the high scale. Hence, are said to be positively distributed. Figure 2.3b confirms that Kenya's population seems to be increasing from 2002-2009. The outlier in each year corresponds to Nairobi with corresponding values shown in Table 2.5 below.

Table 2.5: Counties With Extreme Population For 2002-2009

Year	2002	2003	2004	2005	2006	2007	2008	2009
Nairobi	2,495,170	2,600,859	2,710,706	2,815,838	2,924,309	3,034,379	3,146,303	3,260,124

We now describe risk estimation in the study population in the next section.

2.3 Crude Estimation of Tuberculosis Risk

In this section, we describe how expected number and standardized mortality ratio (SMR) of TB cases were obtained.

2.3.1 Estimates for Spatial Data

Suppose that the unknown risk of TB in region i is given as $\vartheta_i, i = 1, 2, \ldots, n$. 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 $r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} N_i}$ is the overall disease risk in the study population.

We estimate ϑ_i within the frequentist paradigm. We first assume that $y_i \sim Poisson\left(E_i\vartheta_i\right)$. Based on the sample $\mathbf{y} = \{y_1, \ldots, y_n\}$, the likelihood function and the corresponding log-likelihood function are expressed as

$$\ell\left(\vartheta_i\right) = \prod_{i=1}^{n} \frac{\exp\left(-E_i\vartheta_i\right)\left(E_i\vartheta_i\right)^{y_i}}{y_i!} = P\left(\mathbf{y}, \mathbf{E} \mid \boldsymbol{\vartheta}\right) \tag{2.3.1}$$

and

$$\ln \ell\left(\vartheta_i\right) = -\sum_{i=1}^{n} E_i\vartheta_i + \sum_{i=1}^{n} y_i \ln\left(E_i\vartheta_i\right) - \prod_{i=1}^{n} y_i! \tag{2.3.2}$$

The maximum likelihood estimator $\widehat{\vartheta}_i$ of ϑ_i is obtained via $\frac{\partial(\ln \ell(\vartheta_i))}{\partial \vartheta_i} = 0$ and is given by $\widehat{\vartheta}_i = \frac{y_i}{E_i}$. This estimator, $\widehat{\vartheta}_i$ is referred to as the standardized mortality ratio in region i.

The standard error of $\widehat{\vartheta}_i$ is given by $SE\left(\widehat{\vartheta}_i\right) = \frac{\sqrt{y_i}}{E_i}$ and the corresponding $100\left(1-\alpha\right)\%$ Confidence Interval (CI) for $\widehat{\vartheta}_i$ is given by $\widehat{\vartheta}_i \pm Z_{\frac{\alpha}{2}} SE\left(\widehat{\vartheta}_i\right)$.

2.3.2 Estimates for Spatio-Temporal Data

Suppose that the unknown risk of TB in region i and period t is given as $\vartheta_{it}, i = 1, 2, \ldots, n, t = 1, 2, \ldots, T$. Let y_{it} and N_{it} denote the number of TB cases and the population at risk respectively

in region i and period t. The expected number of TB cases in region i and period t can then be written as $E_{it} = rN_{it}$, where $r = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} N_{it}}$ is the overall disease risk in the study population.

We estimate ϑ_{it} within the frequentist paradigm. We first assume that $y_{it} \sim Poisson\left(E_{it}\vartheta_{it}\right)$. Based on y_{it} sample, the likelihood function and the corresponding log-likelihood function are expressed as

$$\ell\left(\vartheta_{it}\right) = \prod_{i=1}^{n}\prod_{t=1}^{T} \frac{\exp\left(-E_{it}\vartheta_{it}\right)\left(E_{it}\vartheta_{it}\right)^{y_{it}}}{y_{it}!} = P\left(\mathbf{y}, \mathbf{E} \mid \boldsymbol{\vartheta}\right) \tag{2.3.3}$$

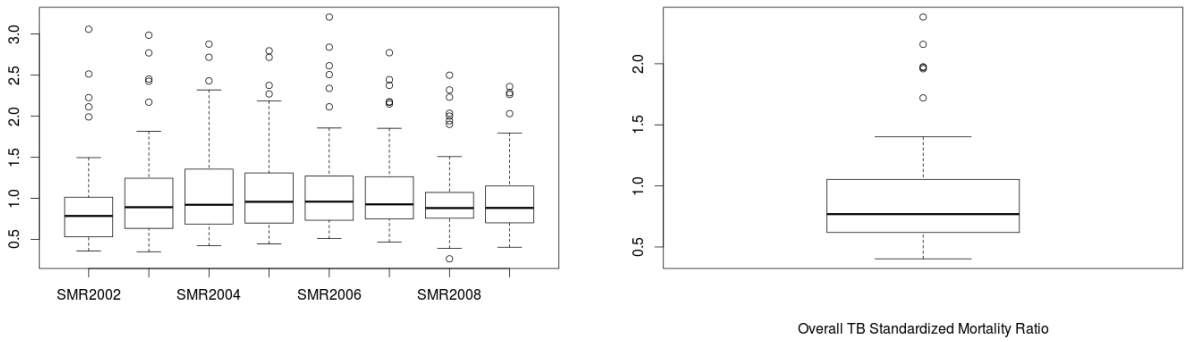
and

$$\ln \ell\left(\vartheta_{it}\right) = -\sum_{i=1}^{n}\sum_{t=1}^{T} E_{it}\vartheta_{it} + \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it} \ln\left(E_{it}\vartheta_{it}\right) - \prod_{i=1}^{n}\prod_{t=1}^{T} y_{it}! \tag{2.3.4}$$

The maximum likelihood estimator $\widehat{\vartheta}_{it}$ of ϑ_{it} is obtained via $\frac{\partial(\ln \ell(\vartheta_{it}))}{\partial \vartheta_{it}} = 0$ and is given by $\widehat{\vartheta}_{it} = \frac{y_{it}}{E_{it}}$. This estimator, $\widehat{\vartheta}_{it}$ is referred to as the standardized mortality ratio in region i and period t.

The standard error of $\widehat{\vartheta}_{it}$ is given by $SE\left(\widehat{\vartheta_{it}}\right) = \frac{\sqrt{y_{it}}}{E_{it}}$ and the corresponding $100\left(1-\alpha\right)\%$ Confidence Interval (CI) for $\widehat{\vartheta}_{it}$ is given by $\widehat{\vartheta}_{it} \pm Z_{-\frac{\alpha}{2}} SE\left(\widehat{\vartheta}_{it}\right)$.

We now present exploratory analysis of the standardized mortality ratio (SMR).



(a) (b)

Figure 2.4: Box plot of County Level TB Standardized Mortality Ratio for 2002-2009

Figure 2.4 enables us to deduce from counties that are likely to exhibit high risk of TB infection for 2002-2009. Distribution of SMR is positively skewed over the study period with high extreme

values and a very low extreme value in 2008. Counties that correspond to highest and lowest risk are summarised in Table 2.6.

Table 2.6: Summary of County Level TB Standardized Mortality Ratio for 2002-2009

Year	Highest Risk		Second Highest Risk		Lowest Risk		Second Lowest Risk	
2002	Mombasa	3.057(2.979,3.135)	Nairobi	2.513 (2.474,2.552)	Vihiga	0.3599 (0.329, 0.391)	Laikipia	0.3691 (0.330, 0.408)
2003	Mombasa	2.985(2.909,3.061)	Nairobi	2.770 (2.730, 2.810)	Nyamira	0.349 (0.319,0,380)	Laikipia	0.406 (0.366, 0.446)
2004	Nairobi	2.877 (2.837, 2.917)	Mombasa	2.717 (2.643,2.791)	Vihiga	0.422 (0.389,0.455)	Laikipia	0.448 (0.406,0.490)
2005	Mombasa	2.795 (2.723,2.867)	Nairobi	2.716 (2.678,2.754)	ElgyeyoMarakwet	0.445 (0.402,0.488)	Vihiga	0.450 (0.417, 0.483)
2006	Marsabit	3.208 (3.058, 3.358)	Mombasa	2.831 (2.760, 2.902)	Tana River	0.509 (0.450,0.568)	ElgyeyoMarakwet	0.539 (0.493,0.586)
2007	Mombasa	2.772 (2.703,2.841)	Nairobi	2.444 (2.409,2.479)	Nandi	0.468 (0.438,0.498)	Tana River	0.511 (0.452,0.569)
2008	Mombasa	2.499 (2.434,2.564)	Nairobi	2.319 (2.285,2.352)	Nandi	0.263 (0.241,0.285)	Nyamira	0.390 (0.360, 0.421)
2009	Mombasa	2.360 (2.298,2.422)	Nairobi	2.285 (2.253,2.318)	Nandi	0.405 (0.378,0.432)	Nyamira	0.473 (0.440,0.506)
Overall	Mombasa	2.383 (2.361, 2.405)	Nairobi	2.159 (2.148, 2.170)	Nandi	0.403 (0.394,0.412)	ElgyeyoMarakwet	0.422 (0.408, 0.446)

Table 2.6 shows counties with the highest and lowest TB SMR with their corresponding 95% confidence interval (CI). County's standardized mortality ratio greater than 1 is considered a high risk county and less than 1 is considered a low risk county. Table 2.6 shows that Mombasa dominates as the highest TB risk county followed by Nairobi over the study period. Their overall SMR are estimated at $2.383(2.361, 2.405)$ and $2.159(2.148, 2.170)$ respectively. Nandi dominates as the lowest TB risk county followed by Laikipia county over the study period. However, the overall SMR shows that Nandi has the lowest risk followed by Elgyeyo Markwet with SMR estimates at $0.403(0.394, 0.412)$ and $0.422(0.408, 0.446)$ respectively.

Table 2.7: Summary of Period Specific TB Standardized Mortality Ratio

Year	Mean	SD	Q_1	Q_2	Q_3	Min	Max
2002	0.9266	0.5850116	0.5305	0.7849	1.0135	0.3599	3.0573
2003	1.0396	0.6186062	0.6339	0.8910	1.2441	0.3492	2.9855
2004	1.1148	0.6064510	0.6867	0.9211	1.3547	0.4225	2.8766
2005	1.1182	0.6042379	0.6984	0.9572	1.3083	0.445	2.7952
2006	1.1818	0.6465715	0.7329	0.9590	1.2712	0.5091	3.2068
2007	1.1274	0.5638802	0.7480	0.9258	1.2627	0.4681	2.7718
2008	1.0387	0.518773	0.7575	0.8812	1.0709	0.2633	2.4985
2009	1.0214	0.498242	0.7012	0.8826	1.1496	0.4046	2.3604
Overall	0.9335	0.4880026	0.6185	0.7690	1.0540	0.4031	2.3830

Table 2.7 presents summary statistics of the standardized mortality ratio for each period, where Q_1, Q_2, and Q_3 represent the first quartile, median, and third quartile respectively. Low risk of TB

prevalence is observed in 2002 but risk increased gradually from 2003-2006 and then decreased slightly from 2007-2009. The overall risk effect is estimated at 0.93 (95% credible interval = 0.62-0.77).

Chapter 3

The Bayesian Hierarchical Modelling

This chapter introduces the Bayesian methods and the computational methodologies based on which parameter estimates in this study are obtained. The development of Bayesian inference has the data likelihood as a fundamental concept [Lawson, 2008]. According to Lawson [2008], the *likelihood principle*, by which observations come to play through the likelihood function, is a fundamental part of the Bayesian paradigm. This implies that information concerning the data is entirely expressed by the likelihood function [Lawson, 2008].

3.1 The Likelihood Function

Let $y_i, i = 1, \ldots, n$ be a random variable with probability density function $P(y_i \mid \boldsymbol{\vartheta})$, where $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_p)$ is a vector of relative risk parameters. The likelihood function of y_i is defined as

$$P(\mathbf{y} \mid \boldsymbol{\vartheta}) = \prod_{i=1}^{n} P(y_i \mid \boldsymbol{\vartheta}). \tag{3.1.1}$$

Equation (3.1.1) is based on the assumption that the sample values of $\mathbf{y} = (y_1, \ldots, y_n)'$ given the parameters $\boldsymbol{\vartheta}$ are independent [Lawson, 2008]. This independence assumption makes it possible for us to express the likelihood function as a product of the individual contributions of $P(y_i \mid \boldsymbol{\vartheta})$ in (3.1.1). Hence, the data is said to be conditionally independent [Lawson, 2008].

In the next section, we discuss prior information about the parameters of interest.

3.2 The Prior Distribution

Bayesian methods are based on prior belief about the parameters of interest. This belief about a parameter is captured in a density function referred to as a prior distribution. The linking of the prior information with data as given in the likelihood leads to posterior inference. We discussed posterior distribution in section 3.3.

A prior distribution is a distribution assigned to the parameter $\boldsymbol{\vartheta}$ before the data y_i are observed [Lawson, 2008]. All parameters within the Bayesian models are stochastic and assigned appropriate prior distribution [Lawson, 2008]. According to Lawson [2008], prior distributions provide additional "data" for a problem, hence can be used to enhance estimation of parameters.

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 $P(\boldsymbol{\vartheta})$. Let us now consider some properties or types of prior distributions.

3.2.1 The Propriety

Inpropriety of prior distribution is the condition where integration of the prior distribution of a random variable $\boldsymbol{\vartheta}$ over its range Ω is infinity [Lawson, 2008]. That is,

$$\int_{\Omega} P(\boldsymbol{\vartheta})\, \mathrm{d}\boldsymbol{\vartheta} = \infty. \tag{3.2.1}$$

A prior distribution is proper if its normalizing constant (See Section 3.3) is infinite [Lawson, 2008]. Lawson [2008] noted that, though impropriety is a limitation of any prior distribution, it is not necessarily the case that an improper prior will lead to impropriety in the posterior.

3.2.2 Conjugate Priors

Sometimes, a particular combination of the prior distributions and the likelihood function lead to the same distribution family in the posterior, where the posterior distribution has distribution

family as that of the prior (See Section 3.3). This type of priors are referred to as the conjugate priors. Conjugates are useful when evaluating complex distributions.

Specifically, for Poisson likelihood with parameter ϑ and the Gamma prior distribution for ϑ, the posterior distribution of ϑ is also Gamma distribution (See Section 3.3, Equation (3.3.5)). Binomial likelihood and beta distribution for the parameter ϑ lead to beta distribution in the posterior [Lawson, 2008]. Similar results hold for normal data likelihood and normal prior distribution.

Conjugacy can be identified by examining the kernel of the prior-likelihood product. The prior-likelihood (unnormalised kernel) should be in a similar form to the prior distribution. According to Lawson [2008], conjugacy guarantees a proper posterior distribution.

3.2.3 Noninformative Priors

Noninformative prior is a type of prior distribution that do not make strong preference over the observed values [Lawson, 2008]. Noninformative prior distributions are sometimes referred to us vague or flat prior distribution. Choosing a noninformative prior distribution for the parameters tends to mean that in any posterior analysis, the prior distribution(s) will have little influence compared to the likelihood of the data.

According to Lawson [2008], the choice of noninformative prior can be made with some general understanding of the range and behaviour of the variables. For instance, variance parameters must have prior distributions on the positive real line. Noninformative priors in this range are often the in the gamma, inverse gamma, or uniform families.

3.2.4 Jeffery's Priors

In Bayesian probability, the Jefferys prior is a non-informative prior distribution on parameter space $\boldsymbol{\vartheta}$ that is proportional to the square root of the determinant of the Fisher information

$I(\boldsymbol{\vartheta})$. Therefore, Jefferys prior for a single parameter ϑ is defined by

$$P(\vartheta) \propto \sqrt{I(\vartheta)}, \quad \text{where} \quad I(\vartheta) = -E\left[\frac{d^2 \ln P(\mathbf{y} \mid \vartheta)}{\mathrm{d}\vartheta^2}\right]. \tag{3.2.2}$$

Hence, the Jefferys prior can alternatively be defined as

$$P(\vartheta) \propto \sqrt{I(\vartheta)} = \sqrt{\left[\left(\frac{d}{\mathrm{d}\vartheta} \ln P(\mathbf{y} \mid \vartheta)\right)^2\right]}. \tag{3.2.3}$$

Jeffery's priors were developed in an attempt to find such vague or flat prior for a given distribution [Lawson, 2008].

Following equation (3.2.3), the Jeffry's prior for Poissson mean ϑ of the positive real value y is defined by

$$P(\vartheta) = \sqrt{\sum_{x=0}^{+\infty} P(y \mid \vartheta)\left(\frac{y-\vartheta}{\vartheta}\right)^2} = \frac{1}{\sqrt{\vartheta}} \tag{3.2.4}$$

This prior is *improper* and not noninformative [Lawson, 2008].

Also, Jeffrys prior for the normal distribution of the positive real value X with fixed mean ϑ is

$$P(\vartheta) = \sqrt{\int_{-\infty}^{+\infty} P(y \mid \vartheta)\left(\frac{y-\vartheta}{\sigma^2}\right)^2 \mathrm{d}y} = \sqrt{\frac{\sigma^2}{\sigma^4}} \propto 1. \tag{3.2.5}$$

This prior does not depend upon ϑ. The Jeffrys prior for the normal distribution of the positive real value y with standard deviation, $\sigma > 0$ is

$$P(\sigma) = \sqrt{\int_{-\infty}^{+\infty} P(y \mid \vartheta)\left(\frac{(y-\vartheta)^2 - \sigma^2}{\sigma^3}\right)^2 \mathrm{d}y} = \sqrt{\frac{2}{\sigma^2}} \propto \frac{1}{\sigma} \tag{3.2.6}$$

Again, using equation (3.2.3), Jeffrys prior for binomial likelihood with parameter ϑ and sample size n is

$$P(\vartheta) = \sqrt{\frac{n}{\vartheta(1-\vartheta)}} = n^{1/2}\vartheta^{-1/2}(1-\vartheta)^{-1/2}. \tag{3.2.7}$$

The Equation (3.2.7) can be written as $P(\vartheta) \propto \vartheta^{-1/2}(1-\vartheta)^{-1/2}$, where $\vartheta \sim Beta\left(\frac{1}{2}, \frac{1}{2}\right)$.

The form of the data likelihood helps to determine the prior (3.2.2) but not the actual observed data since we are averaging over $\boldsymbol{y}$ in equation (3.2.2).

It is should be noted that it is sometimes imperative to be more informative with the prior distributions if the likelihood function has little information about the identification of the parameters Lawson [2008]. Then, identification can only come from the prior specification [Lawson, 2008]. In this instance, Lawson [2008] defined identifiability as an issue relating to the ability to distinguish between parameters with a parametric model.

3.3 The Posterior Distribution

3.3.1 The General Case

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

$$P\left(\boldsymbol{\vartheta} \mid \mathbf{y}\right) = \frac{P\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right) P\left(\boldsymbol{\vartheta}\right)}{\int\limits_{p} L\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right) P\left(\boldsymbol{\vartheta}\right) d_{\boldsymbol{\vartheta}}}, \tag{3.3.1}$$

where $\int\limits_{p} L\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right) P\left(\boldsymbol{\vartheta}\right) d_{\boldsymbol{\vartheta}}$ is called the normalizing constant.

3.3.2 The Poisson-Gamma Distribution for TB incidence Data

Let y_i and $E_i, i = 1, \ldots, E_n$, denote the observed and expected number of TB cases in county i respectively. We assume that $y_i \sim Poisson\left(E_i\vartheta\right)$, where ϑ is the unknown relative risk. That is, the likelihood function for y_i is given by

$$P\left(\mathbf{y} \mid E_i\vartheta\right) = \frac{\left(E_i\vartheta\right)^{y_i} \exp\left(-E_i\vartheta\right)}{y_i!}, y_i, i, \ldots, n. \tag{3.3.2}$$

We assume that $\vartheta \sim Gamma\,(a,b)$ as a prior distribution, $P\,(\vartheta)$ for ϑ defined alternatively by

$$P\,(\vartheta \mid a,b) = \frac{(\vartheta)^{a-1}}{\Gamma\,(a)\,b^{a}} \exp\,(-\vartheta b)\,, \vartheta, a, b > 0. \tag{3.3.3}$$

Therefore, the posterior distribution for ϑ is given by

$$P\,(\vartheta \mid \mathbf{y}) = \frac{\prod_{i=1}^{N} \frac{(E_i\vartheta)^{y_i} \exp(-E_i\vartheta)}{y_i!} \frac{(\vartheta)^{a-1}}{\Gamma(a)b^{a}} \exp\,(-\vartheta b)}{\int \prod_{i=1}^{N} \frac{(E_i\vartheta)^{y_i} \exp\,(-E_i\vartheta)}{y_i!} \frac{(\vartheta)^{a-1}}{\Gamma\,(a)\,b^{a}} \exp\,(-\vartheta b)\,\mathrm{d}\vartheta}, \vartheta > 0, a > 0, b > 0. \tag{3.3.4}$$

This model specification here is referred to as first level hierarchical model. The posterior distribution can be simplified as

$$P\,(\vartheta^* \mid E_i, a^*, b^*, \mathbf{y}) = \frac{b^{*a^*}}{\Gamma\,(a^*)} \vartheta^{a^*-1} \exp\,(-\vartheta b^*)\,, \tag{3.3.5}$$

where $a^* = \sum_{i=1}^{n} y_i + a$ and $b^* = \sum_{i=1}^{n} E_i + b$. Therefore, the posterior mean of the Poisson-Gamma model given

$$E\,(\vartheta^* \mid E_i, a^*, b^*, \mathbf{y}) = \frac{\sum_{i=1}^{n} X_i + a}{\left(\sum_{i=1}^{n} E_i + b\right)} \tag{3.3.6}$$

and the posterior variance is given by

$$var\,(\vartheta^* \mid E_i, a^*, b^*, \mathbf{y}) = \frac{\sum_{i=1}^{n} y_i + a}{\left(\sum_{i=1}^{n} E_i + b\right)^2} \tag{3.3.7}$$

The posterior mean (3.3.6) can be alternatively expressed as

$$E\,(\vartheta^* \mid y_i, a^*, b^*) = \frac{y_i + a}{E_i + b} = (1 - R_i)\,\widehat{\vartheta}_i + R_i \frac{a}{b}, \quad \text{where} \quad R_i = \frac{b}{(E_i + b)} \tag{3.3.8}$$

The above description of estimation of parameters is referred to us the Empirical Baye Estimation (EBE).

In this study, we employed hierarchical Bayes approach to estimate parameters of interest discussed in Section 4.1. We used Markov Chain Monte Carlo method for simulation of parameters from their respective distributions via Gibbs Sampling discussed in the next section.

3.4 Bayesian Markov Chain Monte Carlo (MCMC) Method.

In this section, we will provide the elementary notion of MCMC algorithm used to carry out posterior inference in the case where the product of the likelihood and the prior are analytically intractable.

Ntzoufras [2011] stated that MCMC methods enabled quantitative researchers to use highly complicated models to estimate the corresponding posterior distribution with accuracy. The development of modern MCMC has greatly contributed to the development of and propagation of Bayesian theory [Ntzoufras, 2011].

The MCMC methods involve construction of a Markov Chain (MC) that eventually "converges" to the target (stationary) distribution [Ntzoufras, 2011]. The target distribution in this thesis is the Posterior distribution $P(\boldsymbol{\vartheta} \mid \mathbf{y})$. In the next Section, we explain how MCMC algorithms work.

3.4.1 Markov Chain Monte Carlo Algorithm

Let $\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)}, \ldots, \boldsymbol{\vartheta}^{(G)}$ be a sample of size G from the posterior distribution $P(\boldsymbol{\vartheta} \mid \mathbf{y})$. A Markov Chain is a stochastic process defined by $\{\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)}, \ldots, \boldsymbol{\vartheta}^{(G)}\}$ such that $P\left(\boldsymbol{\vartheta}^{(g+1)} \mid \boldsymbol{\vartheta}^{(g)}, \ldots, \boldsymbol{\vartheta}^{(1)}\right) = P\left(\boldsymbol{\vartheta}^{(g+1)} \mid \boldsymbol{\vartheta}^{(g)}\right)$. That is, the distribution of $\boldsymbol{\vartheta}$ at time $g+1$ given all the preceding $\boldsymbol{\vartheta}$ values (for $g, g-1, \ldots, 1$) depends only on the value $\boldsymbol{\vartheta}^{(g)}$ of the previous sequence g.

As $g \to \infty$, the distribution $\boldsymbol{\vartheta}^{(g)}$ converges to its equilibrium, which is independent of the initial value of the chain $\boldsymbol{\vartheta}^{(0)}$ [Ntzoufras, 2011]. This condition occurs when the MC is irreducible, aperiodic, and positive-recurrent [Nummelin, 2004].

We describe an MCMC algorithm for which this property holds. The standard approach to Bayesian inference using MCMC is as follows:

1. Select an initial value $\boldsymbol{\vartheta}^{(0)}$.

2. Generate G values until the equilibrium distribution is reached.

3. Monitor the converge of the algorithm using the convergence diagnostic. If convergence

diagnostics fail, we generate more samples.

4. Cut off the first B observations. This is referred to as the Burn-in period. Here, the first B iterations are eliminated from the sample to avoid the influence of initial values. Ntzoufras [2011] stated that if the sample generated is large enough, burn-in period's effect on the calculation of the posterior is minimal.

5. Consider $\{\boldsymbol{\vartheta}^{(B+1)}, \boldsymbol{\vartheta}^{(B+2)}, \ldots, \boldsymbol{\vartheta}^{(G)}\}$ as the sample for the posterior analysis.

6. Plot the posterior distribution.

7. Finally, obtain summaries of the posterior distributions.

It is items $6-7$ in the list above that we refer to the *convergence diagnostics*. Convergence diagnostics is used to identify cases where convergence is not achieved.

The MCMC output provides us with a random sample of the form $\{\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)}, \ldots, \boldsymbol{\vartheta}^{(G')}\}$. From this sample, for any function $M(\boldsymbol{\vartheta})$ of parameters $\boldsymbol{\vartheta}$ of interest, we can obtain a sample of the desire parameter $M(\boldsymbol{\vartheta})$ by considering $M\left(\boldsymbol{\vartheta}^{(1)}\right), M\left(\boldsymbol{\vartheta}^{(2)}\right), \ldots, M\left(\boldsymbol{\vartheta}^{(G')}\right)$. We can also obtain any posterior summary of $M(\boldsymbol{\vartheta})$ from the sample distribution using the traditional sample estimates [Ntzoufras, 2011]. Thus we can estimate the posterior mean $\widehat{E}(M(\boldsymbol{\vartheta}) \mid \mathbf{y})$ by

$$\widehat{E}(M(\boldsymbol{\vartheta}) \mid \mathbf{y}) = \frac{1}{G'} \sum_{g=1}^{G'} M(\boldsymbol{\vartheta}^g) \tag{3.4.1}$$

and the posterior standard deviation $\widehat{SD}$ by

$$\widehat{SD}(M(\boldsymbol{\vartheta}) \mid \mathbf{y}) = \frac{1}{G'} \sum_{g=1}^{G'-1} \left[M(\boldsymbol{\vartheta}^g) - \widehat{E}(M(\boldsymbol{\vartheta}) \mid \mathbf{y})\right]^2 \tag{3.4.2}$$

Other measures of interest such as the posterior median or quantiles can be obtained in similar fashion [Ntzoufras, 2011].

According to Ntzoufras [2011], the two most popular MCMC methods are the Metropolis-Hasting algorithm (Metropolis et al 1953; Hastings, 1970) and the Gibbs sampler (Geman and Geman 1984). Other MCMC algorithms that appeared in recent MCMC literature are the slice sampler

(Higdon, 1998; Damien et al., 1999; Neal, 2003), the reversible jump MCM (RJMCMC) algorithm (Green, 1995) and the perfect sampling (Propp and Wilson, 1996; Merller, 1999) [Ntzoufras, 2011].

In the next section, we focus on the Gibbs sampler as it is the MCMC implemented in WinBUGS.

3.4.2 The Gibbs Sampler (GS)

The Gibbs sampler was introduced by Geman and Geman (1984) as an MCMC algorithm for simulating samples from the posterior distribution [Ntzoufras, 2011]. Algorithm of Gibbs sampling is summarized below:

1. Set initial values $\boldsymbol{\vartheta}^{(0)}$.
2. For $g = 1, 2, \ldots, G$, repeat the steps below:
   - Set $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^{(g-1)}$.
   - For $r = 1, \ldots, d$, update ϑ_r from $\vartheta_r \sim P\left(\vartheta_r \mid \boldsymbol{\vartheta}_{/r}, \mathbf{y}\right)$, where

$$\boldsymbol{\vartheta}_{/r} = (\vartheta_1, \ldots, \vartheta_{r-1}, \vartheta_{r+1}, \ldots, \vartheta_d)\,. \tag{3.4.3}$$

   - Set $\boldsymbol{\vartheta}^{(g)} = \boldsymbol{\vartheta}$ and save it as generated set of values at $g+1$ iteration of the algorithm.

Hence, given a particular state of the chain $\boldsymbol{\vartheta}^{(g)}$, we generate new values by

$$\begin{array}{lll}
\vartheta_1^{(g)} & from & P\left(\vartheta_1 \mid \vartheta_2^{(g-1)}, \ldots, \vartheta_i^{(g-1)}, \mathbf{y}\right) \\
\vartheta_2^{(g)} & from & P\left(\vartheta_2 \mid \vartheta_1^{(g)}, \vartheta_3^{(g-1)}, \ldots, \vartheta_i^{(g-1)}, \mathbf{y}\right) \\
\vartheta_3^{(g)} & from & P\left(\vartheta_3 \mid \vartheta_1^{(g)}, \vartheta_2^{(g)}, \vartheta_4^{(g-1)}, \ldots, \vartheta_i^{(g-1)}, \mathbf{y}\right) \\
\vdots & \vdots & \vdots \\
\vartheta_r^{(g)} & from & P\left(\vartheta_r \mid \vartheta_1^{(g)}, \vartheta_2^{(g)}, \ldots, \vartheta_{r-1}^{(g)}, \vartheta_{r+1}^{(g-1)}, \ldots, \vartheta_i^{(g-1)}, \mathbf{y}\right) \\
\vdots & \vdots & \vdots \\
\vartheta_i^{(g)} & from & P\left(\vartheta_i \mid \vartheta_1^{(g)}, \vartheta_2^{(g)}, \ldots, \vartheta_{i-1}^{(g)}, \mathbf{y}\right).
\end{array}$$

Ntzoufras [2011] noted that generating values from

$$P\left(\vartheta_r \mid \boldsymbol{\vartheta}_{/r}, \mathbf{y}\right) = P\left(\vartheta_r \mid \vartheta_1^{(g)}, \ldots, \vartheta_{r-1}^{(g)}, \vartheta_{r+1}^{(g-1)}, \ldots, \vartheta_i^{(g-1)}, \mathbf{y}\right)' \tag{3.4.4}$$

is relatively easy since it is a univariate distribution and can be written as $P\left(\vartheta_r \mid \boldsymbol{\vartheta}_{/r}, \mathbf{y}\right) \propto P\left(\boldsymbol{\vartheta} \mid \mathbf{y}\right)$, where all the variables except ϑ_r are kept constant at their given values.

3.4.3 Assessing and Improving Markov Chain Monte Carlo Convergence

It is important to decide how many iterations to use to represent the posterior density and to ensure that the Markov chain converged. It should be noted that convergence of a model does not necessarily imply a good model. It is just the beginning of model assessment.

- **Autocorrelation Function (ACF) Plots**: Congdon [2010] stated that nonvanishing autocorrelation at high lags indicates less information about the posterior is provided by each iterate and a high sample size is required to cover the parameter space. Autocorrelation is a situation where parameters in the chain are correlated. Autocorrelation can be reduced by "thinning". Thinning involves storing of samples from every k^{th} iteration, where $k > 1$ is the value of the field thinned [Congdon, 2010]. Thinning reduces MCMC error and storage requirements especially when long runs are being carried out [Lawson et al., 2003].

  Another way of reducing correlation within the chain is the use of "over-relax" algorithm. This generates multiple samples at each iteration and then select one that is negatively correlated to the current value [Lawson et al., 2003].

- **Kernel Density Plots**: A more satisfactory density plot for a converged chain would look more bell-shaped or parameters whose marginal posterior densities are approximately normal [Lawson et al., 2003].

- **Gelman and Rubin Multiple Chain Convergence**: Gelman and Rubin multiple chain convergence diagnostics is based on using two or more parallel chain with divers starting values [Lawson et al., 2003]. Lawson et al. [2003] stated that multiple chain convergence diagnostics provide evidence for the robustness of convergence across different subspace.

Convergence of a Markov Chain can be improved by standardizing covariates and the unstructured random effect (See Section 4.2) [Congdon, 2010].

3.4.4 Criteria for Model Selection

In this section, we describe one of the methods used to select the best fitting model from a set candidate models. Though technology to fit complex models through the Bayesian hierarchical models is widely available, there is no clear criteria to compare models and select best models. The most widely used criteria is how to measure and appropriately penalized the complexity of a hierarchical model.

The most commonly and widely used criteria for comparing hierarchical models is the Deviance Information Criterion proposed by Spiegelhalter et al. [2002]. The DIC works in a similar manner like that Bayesian Information Criterion (BIC) [Schwarz, 1978]. The DIC includes terms for both the fit and the complexity of a model. The BIC is defined as

$$\text{BIC} \quad = D\left(\boldsymbol{\vartheta}\right) + p \log n, \tag{3.4.5}$$

where p is the number of parameters, n is the number of observations and

$$D\left(\boldsymbol{\vartheta}\right) = -2 \log P\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right) \quad \text{is the deviance with parameter vector} \quad \boldsymbol{\vartheta}. \tag{3.4.6}$$

The $D\left(\boldsymbol{\vartheta}\right)$ ignores the standardization term which does not affect the model comparison. The deviance is approximated using the plug-in estimate of the parameter vector $\boldsymbol{\vartheta}$. One similar criterion for models comparison is the Akaike Information Criterion Akaike [1981], defined as

$$\text{AIC} \quad = D\left(\boldsymbol{\vartheta}\right) + 2p. \tag{3.4.7}$$

It should be noted that the researcher is interested in the change of AIC or BIC between two models. These criteria are linear functions of p and D. Limitation of the AIC and the BIC is that they cannot be applied directly to hierarchical models since it is not clear how to defined p in the model.

Spiegelhalter et al. [2002] proposed to estimate p. Given the likelihood function, $P\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right)$, the deviance is usually defined as $D\left(\boldsymbol{\vartheta}\right) = -2 \ln P\left(\mathbf{y} \mid \boldsymbol{\vartheta}\right)$ and the posterior average deviance

$E_{\boldsymbol{\vartheta}|\mathbf{y}}(D)$ or $\bar{D}$ is defined as $\bar{D} = -\frac{2}{G}\sum_{g=1}^{G} P\left(\mathbf{y} \mid \boldsymbol{\vartheta}^g\right)$, where $\boldsymbol{\vartheta}^g$ is a sample parameter value. The deviance of the posterior expected parameter estimate, $\boldsymbol{\vartheta}$ is defined as $\widehat{D}\left(\widehat{\boldsymbol{\vartheta}}\right) = -2\ln P\left(\mathbf{y} \mid \widehat{\boldsymbol{\vartheta}}\right)$. That is, given any sample parameter value $\boldsymbol{\vartheta}^g$, the deviance of the posterior expected parameter estimate is defined as $\widehat{D}\left(\boldsymbol{\vartheta}^g\right) = -2\ln P\left(\mathbf{y} \mid \boldsymbol{\vartheta}^g\right)$. Spiegelhalter et al. [2002] therefore defined the effective number of parameters, pD identified by a model as

$$pD = \bar{D} - D\left(\widehat{\boldsymbol{\vartheta}}\right), \tag{3.4.8}$$

where $D\left(\widehat{\boldsymbol{\vartheta}}\right)$ is the deviance evaluated at the posterior mean of the parameters. Hence, the Deviance Information Criterion (DIC) is defined as

$$\text{DIC} = pD + \widehat{D} \quad \text{or} \quad \text{DIC} \quad = 2\bar{D}\left(\boldsymbol{\vartheta}\right) - D\left(\bar{\boldsymbol{\vartheta}}\right). \tag{3.4.9}$$

For non-hierarchical models, the DIC is seen as a generalization of the Akaike's Criterion (AIC), where $DIC \approx AIC$ [Best, 2011]. The DIC balances the requirement between models fit and low complexities; models fit improves as more parameters are added to the model [Gimenez et al., 2009].

Considering a hierarchical analysis of variance model (ANOVA) model; $y_i \mid \mu_i \sim N\left(\mu_i, \sigma^{-1}\right), \mu_i \sim N\left(\xi, \zeta^{-1}\right), i = 1, \ldots, p$, Spiegelhalter and Coworkers showed that $pD = \sum_i \frac{\tau_i}{\tau_i + \zeta}$. They also defined defined pD alternatively as $pD = \sum_i \Upsilon_i$, where Υ_i is the intraclass correlation coefficient for group i. They noted that if $\tau_i >> \zeta$ for all groups; that is $\Upsilon_i \approx 1$, then $pD \approx p$.

As discussed above, the DIC proposed by Spiegelhalter et al. [2002] posed numerous criticisms. One limitation of the DIC is that it depends on the parameterization of $\boldsymbol{\vartheta}$. Their original paper showed that the degree of DIC dependence on $\boldsymbol{\vartheta}$ varies according to the model. Another limitation of the DIC is the one noted by Spiegelhalter et al. [2002], "Given the number of approximations and assumptions that are required to obtained the DIC, it can only really be used as a broad brush technique for discriminations between obviously disparate models, in much the same way any of the alternative information criterion suggested by BIC and AIC might be used"

Utmost care must be taken when computing pD since it can be negative due to $D\left(\widehat{\boldsymbol{\vartheta}}\right) > \bar{D}$. Instability in the estimate $\widehat{pD}$ of pD can results in limited use of this DIC. This situation most

often occurs in mixture models, models with multiple modes due to over dispersion, inappropriate choice of hyper-parameters for the variance parameters in the hierarchical models, and also non-linear transformation [Lawson et al., 2003].

Having identified the best fitting model using the DIC and pD, it is necessary to investigate the presence of clustering of risk among regions or counties.

The next section presents a brief description of clustering and clusters of risk and clustering and clusters detection methods.

3.5 Disease Cluster Detection

This section presents information about clustering and clusters of disease. The section also discussed methods used in disease mapping to detect elevated risk. Cluster detection is focused on local features of risk surface where elevated or depressed risk of disease occur [Lawson, 2008].

According to Pfeiffer et al. [2008], the term "clustering" is use to describe spatial aggregation of disease events and Wakefield and Waller [2000] stated that a disease is said to be clustered if there is "residual spatial variation in risk after covariate influences have been accounted for [Lawson, 2008]. There exist a variety of cluster and clustering definitions. However, care must be taken when defining cluster and clustering since any difference in their definitions will lead to difference in the ability to detect clusters and clustering detection [Lawson, 2008].

Besag and Newell [1991] classified two different methods of analysing clusters as either "global" or "local". Typical example of global clustering is the correlated heterogeneity term in the Bessag York Model (See Section 5.2) [Lawson, 2008]. They referred to global clustering methods as methods used to assess whether clustering is apparent throughout the study region. Local clustering methods define the locations and extend of clustering [Pfeiffer et al., 2008]. It should be noted that the term clustering is applied to the global clustering methods of cluster analysis, while cluster detection refers to the local methods [Pfeiffer et al., 2008].

Clustering of disease can occur due to various reasons such as: the infectious spread of disease, the occurrence of disease vector in a specific locations and the existence of potential health hazards such as localized pollution sources scattered throughout the region. Each of this creates an increased of risk of disease in its immediate vicinity [Pfeiffer et al., 2008].

Lawson [2008] noted that relative risk (RR) should not be taken for cluster detection since RR estimation concerns the global smoothing of risk and estimation of true underlying risk level. However, the difference between cluster detection and RR is likely to be blurred since methods that are used for RR estimation can be extended to allow for certain types clustering detection (as in Section 5.2).

Having understood the nature of clustering and clusters that may occur in disease mapping, we

present one of the methods used for cluster detection that uses the posterior $P(\boldsymbol{\vartheta} \mid \mathbf{y})$ measures.

3.5.1 Exceedence Probability

The exceedence probability is one of the cluster detection methods that relies on the posterior $P(\boldsymbol{\vartheta} \mid \mathbf{y})$ measures for cluster detection [Lawson, 2008]. Whenever the tendency of clustering in risk is suspected in these estimates, we examine their posterior sample behaviour [Lawson et al., 2003].

The most commonly criteria for cluster detection method is the exceedence probability in relation with the relative estimates for individual areas or counties [Wakefield and Waller, 2000]. The exceedence probability is defined as the probability that the relative risk $\boldsymbol{\vartheta}$ exceeds some threshold level $\Bbbk$, defined by $P(\boldsymbol{\vartheta} > \Bbbk)$. Exceedence probability is computed from the posterior sample values $\{\boldsymbol{\vartheta}_i^g\}_{\{}g = 1, 2 \ldots, G\}$ through $\widehat{P}(\boldsymbol{\vartheta}_i > \Bbbk) = \frac{1}{G} \sum_{g=1}^{G} I(\boldsymbol{\vartheta}_i^g > \Bbbk)$, where

$$I(\boldsymbol{\vartheta}_i^g > \Bbbk) = \begin{cases} 1 & \text{if} \quad \boldsymbol{\vartheta}_i^g > \Bbbk \\ 0 & \text{Otherwise.} \end{cases} \tag{3.5.1}$$

In evaluating $P(\boldsymbol{\vartheta}_i > \Bbbk)$, $\Bbbk$ and the threshold probability must be chosen such that $P(\boldsymbol{\vartheta}_i > \Bbbk) > k$. By convention, k can take the values of $0.95, 0.975, 0.99$, etc. [Lawson et al., 2003].

According to Lawson [2008], exceedence probability is only capable of detecting *hot* spot cluster and does not consider any other information concerning possible forms of cluster or even neighbourhood information. Lawson [2008] defined *Hot* spot as any area displaying "excess" or "unusual" risk.

According Lawson [2008], Hossain and Lawson have some attempts to enhance the exceedence probability by inclusion of neighbourhoods. They stated that, for the neighbourhood of the i^{th} area defined as a_i and the number of neighbours as n_i, then

$$\bar{A}_i = \frac{\sum_{j^*=0}^{n_i} A_{ij^*}}{(n_i + 1)} \quad \text{where} \quad A_{ij^*} = P(\boldsymbol{\vartheta}_{j^*} > \Bbbk) \quad \forall j^* \in a_i \quad \text{and} \quad A_{i0} = P(\boldsymbol{\vartheta}_i > \Bbbk) \tag{3.5.2}$$

These measures can be used to detect different forms of clustering [Lawson, 2008]. However, the usefulness of the exceeedence 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 [Lawson, 2008].

In the next chapter, we present nonspatial models used in disease mapping.

Chapter 4

Uncorrelated Heterogeneity Methods for Relative Risk Estimation

This chapter presents two nonspatial models used in disease modelling and mapping. These are the Poisson-Gamma and the Poisson Log-Normal 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.

4.1 The Poisson-Gamma Model (PG)

The Poisson-Gamma model for relative risk estimation is a gamma prior distribution for the relative risk combines with the Poisson likelihood function for the disease counts to give 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 [Lawson et al., 2003].

4.1.1 Model Description

Let y_i and $E_i, i = 1, \ldots, n$, denote the observed and expected number of TB cases in county i. We assume that $y_i \sim Poisson\left(E_i\vartheta\right)$, where ϑ is the unknown relative risk and Poisson mean μ is $\mu_i = E_i\vartheta$. We assume that $\vartheta \sim Gamma\left(a, b\right)$. The likelihood function for y_i is denoted by

$$\ell\left(\vartheta\right) = \prod_{i=1}^{N} \frac{\left(E_i\vartheta\right)^{y_i} \exp\left(-E_i\vartheta\right)}{y_i!} = P\left(\mathbf{y}, \mathbf{E} \mid \vartheta\right). \tag{4.1.1}$$

and prior distribution for ϑ denoted by

$$P(\vartheta \mid a, b) = \frac{(\vartheta)^{a-1}}{\Gamma(a)\, b^{a}} \exp(-\vartheta b), \vartheta, a, b > 0. \tag{4.1.2}$$

4.1.2 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 hyperparameters 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 modelling using the Poisson-Gamma model.

In this thesis, 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

$$P(\vartheta, a, b, \mid \mathbf{y}, \mathbf{E}) \propto P(\mathbf{y}, \mathbf{E} \mid \vartheta, a, b,)\, P(\vartheta)\, P(a \mid \omega)\, P(b \mid \boldsymbol{\phi}). \tag{4.1.3}$$

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 1,000 sample and thinning of every 30^{th} element of the sample. Figure 4.1 presents the MCMC's convergence diagnostics plots.

4.1.3 Markov Chain Monte Carlo Diagnosis

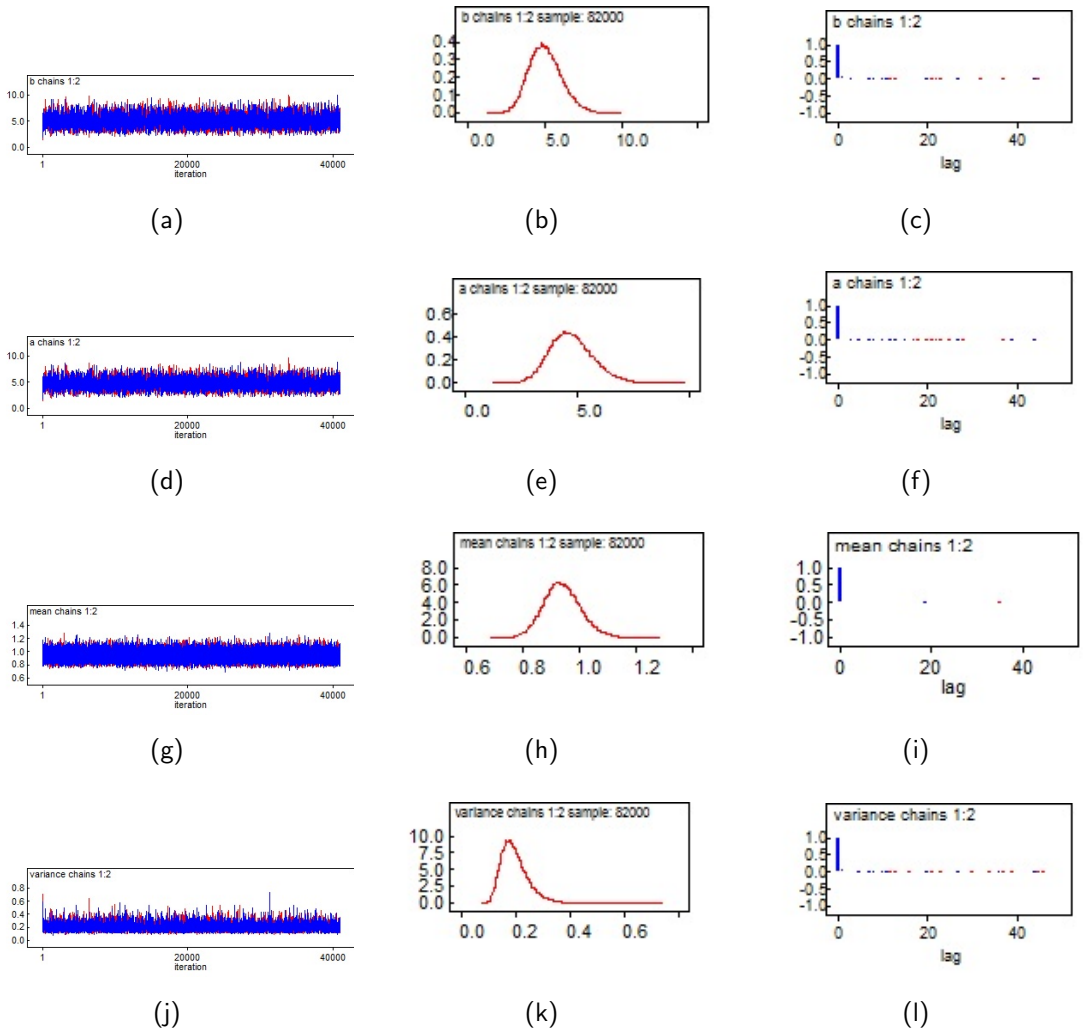


Figure 4.1: Poisson-Gamma Model: Convergence Diagnosis of Markov Chain Monte Carlo.

Figure 4.1 presents Gelman and Rubin convergence diagnostics of the Poisson-Gamma model: Column-wise from the top left, Figure 4.1 (a)-(j) are trace plots for a, b, the mean and variance respectively. Figure 4.1 (b)-(k) are posterior marginal density plots for b, a, the mean and the variance respectively. Figure 4.1 (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. Hence, the density plots for the parameters show that convergence of the chain has reached.

We now present posterior statistics of the Poisson-Gamma model in Table 4.1.

4.1.4 Results of the Poisson-Gamma Model

We now present the results for fitting the Poisson-Gamma model to Kenya TB data. Table 4.1 presents a summary of the Poisson-Gamma model fitted.

Table 4.1: 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	-

From Table 4.1, the mean of 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) in Section 2.3. 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 [Lawson

et al., 2003].

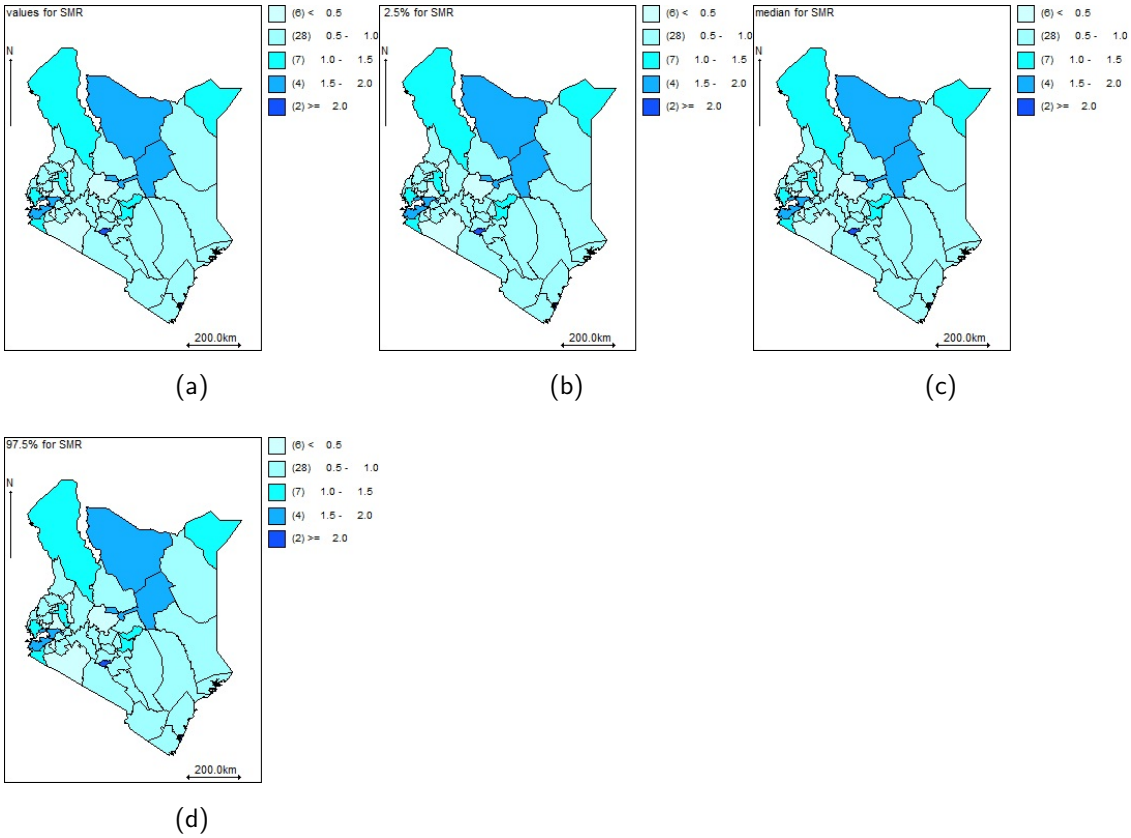


Figure 4.2: Kenya county level Standardized Mortality Ratio's maps: (4.2a) The mean of the SMR and its 2.5% quantile (4.2b), median (4.2c) and 97.5% quantile (4.2d).

Figure 4.2 shows standardized mortality ratio for TB prevalence in the counties of Kenya for 2002-2009. The SMRs vary around their mean, 0.93 with standard deviation, 0.49 (as discussed in Table 4.1). 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).

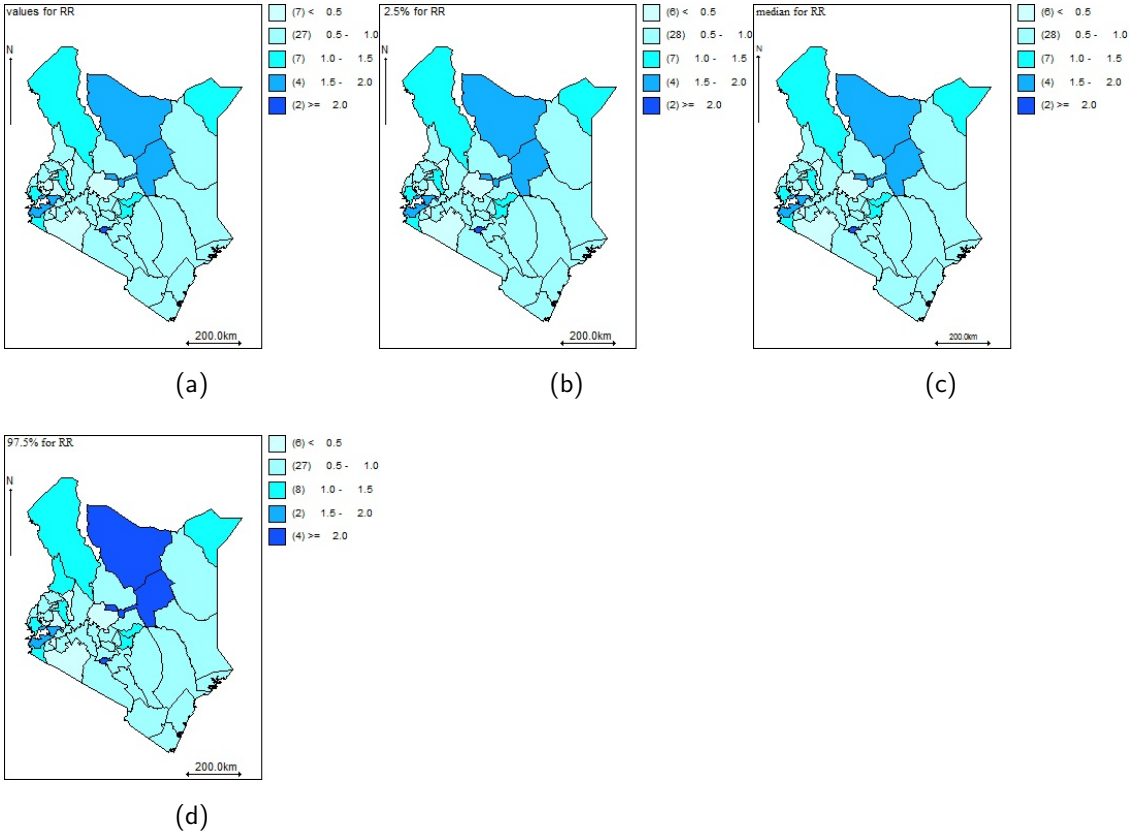


Figure 4.3: Kenya county level Poisson-Gamma posterior mean relative risk maps: (4.3a) The mean of the posterior relative risk and its 2.5% quantile (4.3b), median (4.3c) and 97.5% quantile (4.3d).

From Figure 4.3, 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.

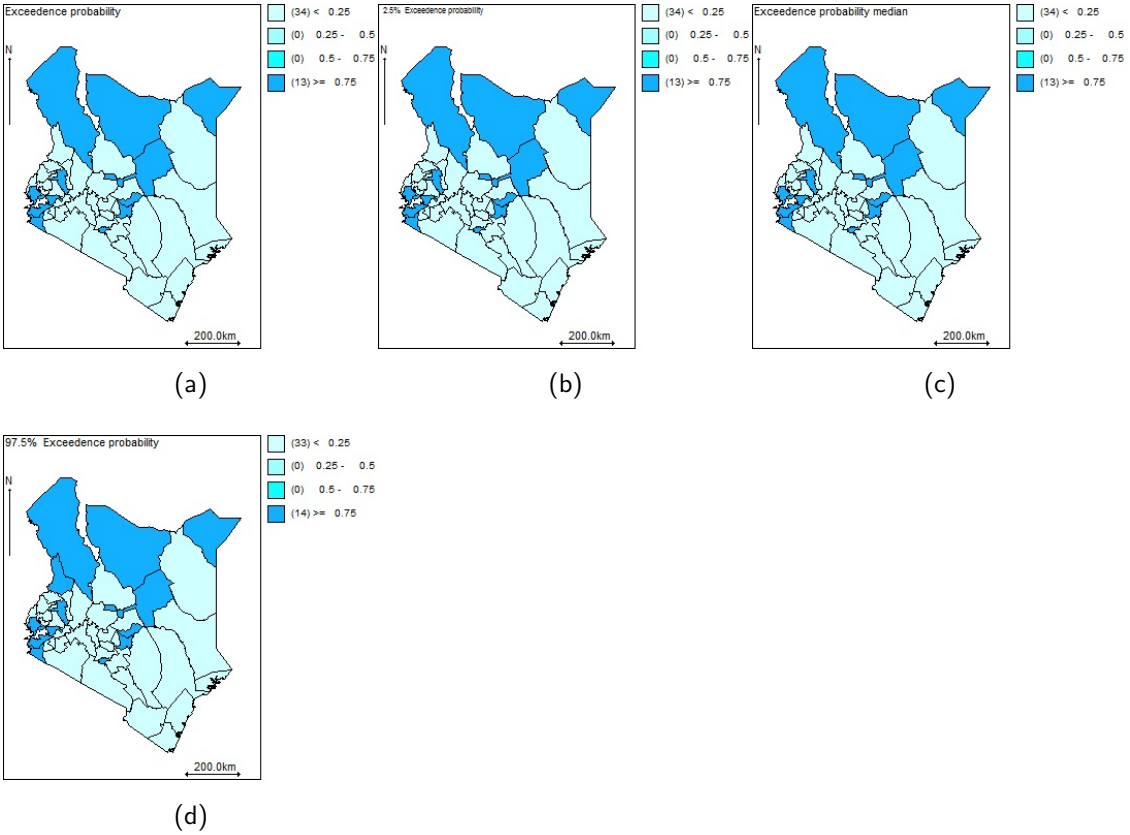


Figure 4.4: Poisson-Gamma posterior relative risk exceedence probability map: Row-wise from the top left Figure: (4.4a) the posterior mean relative risk exceedence probability and its 4.4b 2.5% quantile for relative risk, (4.4c) median for the relative risk and (4.4d) 97.5% quantile for the relative risk.

The map of the exceedence probability in Figure: 4.4 above 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 Gishu, Isiolo, Marsabit, Siaya, Tharaka-Nithi, Mandera, and Embu. This map confirms with Figure 4.3 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 [Lawson et al., 2003]. We therefore present models that nullify theses limitations in the next sections.

4.2 Poisson-Log-Normal Model

We now present a model which allows for flexibility of covariates adjustments or incorporation. Lawson et al. [2003] noted that in disease mapping, the log-normal model is important as it provides a specification that allows for incorporation of covariates.

4.2.1 Model Description

Let y_i and E_i be the observed number and the expected number of disease counts in region $i, i = 1, 2, \ldots, n$ respectively. Further let ϑ_i be the relative risk of disease in region i.

We first consider a situation of a Poisson Log-Normal model with no area-specific random effect u_i and covariate. As stated in the previous section, $y_i \sim Poisson\left(E_i \exp\left(\eta_i\right)\right)$, where $\vartheta_i = \exp\left(\eta_i\right)$ is the exponential of the linear link function and $\mu_i = E_i \exp\left(\eta_i\right)$ is the Poisson mean. Fitting a generalized linear model with a log-link function, we have $\log\left(\mu_i\right) = \log\left(E_i\right) + \eta_i$. By Bayesian paradigm, we assumed that $\eta_i \sim N\left(\mu, \tau^2\right)$ and its hyperparameters, $\mu \sim N\left(0, 1 \times 10^{-6}\right)$ and $\tau^2 \sim Gamma\left(0.5, 0.05\right)$.

Parameter estimation was carried out using Bayesin Markov Chain Monte Carlo via Gibbs Sampling (See Section 3.4). Convergence of the MCMC was reached at 11000 iteration after a burnin period of 10,000 sample and thinning of every 30^{th} element of the chain. Convergence diagnosis plots are presented in Figure 4.5 and posterior statistics of parameters presented in Table 4.2.

We now consider a Poisson Log-Normal 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 $\mathbf{X}$ represents the covariates matrix. The Poisson-Lognormal non-spatial model is given by

$$y_i \mid E_i, X_{ic}, \eta_i \overset{ind}{\sim} Poisson\left(E_i \exp\left(\eta_i\right)\right), \tag{4.2.1}$$

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

$$\log(\vartheta_i) = \eta_i = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ic} + u_i. \tag{4.2.2}$$

From equation (4.2.2), we can write the relative risk as

$$\vartheta_i = \exp\left(\beta_0 + \sum_{p=1}^{P} \beta_p X_{ic}\right), \tag{4.2.3}$$

where ϑ_i are the relative risk of region i, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ 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

$$\log(\mu_i) = \log(E_i) + \beta_0 + \sum_{p=1}^{P} \beta_p X_{ic}. \tag{4.2.4}$$

4.2.2 Parameter Estimation

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

$$\ell(\boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{u}) = \prod_{i=1}^{n} \frac{(E_i \exp(\eta_i))^{y_i} \exp(-E_i \exp(\eta_i))}{y_i!} = P(\mathbf{y}, \mathbf{E} \mid \boldsymbol{\beta}, \boldsymbol{\vartheta}, \boldsymbol{u}), i = 1, 2, \ldots, n. \tag{4.2.5}$$

We would need the prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{u}$ to obtain the posterior distribution for the parameters of interest.

4.2.3 The Likelihood Function of a Regression With Gaussian Random Effect

Consider a response random variable $\mathbf{y} \sim N(0, \boldsymbol{\Sigma})$, 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

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) = N(0, \boldsymbol{\Sigma}), \tag{4.2.6}$$

where $\boldsymbol{X}$ is a design matrix of the covariates. Just to simplify analytic calculation, we can alternatively write the Gaussian linear model equation (4.2.6) as

$$P(\mathbf{y} \mid \boldsymbol{\beta}) = \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{z} - \mathbf{M}\boldsymbol{\beta})^{'}(\boldsymbol{z} - \mathbf{M}\boldsymbol{\beta})\right], \tag{4.2.7}$$

where $z_i = \frac{y_i}{\sigma_i^2}$, $M_{ci} = \frac{X_{ci}}{\sigma_i^2}$, and σ_i^2 are the diagonal covariance matrix $\boldsymbol{\Sigma} = \{\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\}$ with the standard model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$. 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

$$\ln \ell(\boldsymbol{\beta}) = \ln \ell(\boldsymbol{\beta}_{ML}) + \frac{\partial \ln \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mid_{\boldsymbol{\beta}_{ML}} (\boldsymbol{\beta} - \boldsymbol{\beta}_{ML}) + \frac{1}{2}\frac{\partial^2 \ln \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \mid_{\boldsymbol{\beta}_{ML}} (\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})^2, \tag{4.2.8}$$

where $\boldsymbol{\beta}_{ML}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$. Letting $\Omega = \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{\Sigma}|^{1/2}}$ and taking $\log$ of equation (4.2.7), we have

$$\ln P(\mathbf{y} \mid \boldsymbol{\beta}) = \ln \Omega - \frac{1}{2}(\boldsymbol{z}^{'}\boldsymbol{z} - 2\boldsymbol{z}^{'}\mathbf{M}\boldsymbol{\beta} + \boldsymbol{\beta}^{'}\mathbf{M}^{'}\mathbf{M}\boldsymbol{\beta}) \tag{4.2.9}$$

Finding the derivative of equation (4.2.9), it follows that

$$\frac{\partial \ln P(\mathbf{y} \mid \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{z}^{'}\mathbf{M} - \boldsymbol{\beta}^{'}\mathbf{M}^{'}\mathbf{M} = 0 \tag{4.2.10}$$

Solving equation (4.2.10), we have

$$\boldsymbol{\beta}_{ML} = (\mathbf{M}^{'}\mathbf{M})^{-1}\mathbf{M}^{'}\boldsymbol{z}. \tag{4.2.11}$$

Now finding the second derivative of equation (4.2.13), we have:

$$\frac{\partial^2 \ln P(\mathbf{y} \mid \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = -\frac{\partial(\boldsymbol{\beta}^{'}\mathbf{M}^{'}\mathbf{M})}{\partial \boldsymbol{\beta}} = -\mathbf{M}^{'}\mathbf{M} = H. \tag{4.2.12}$$

Hence we can rewrite equation (4.2.10) as

$$\boldsymbol{\beta}_{ML} = (\mathbf{M}^{'}\mathbf{M})^{-1}\mathbf{M}^{'}\boldsymbol{z} = \mathbf{H}^{-1}\mathbf{M}^{'}\boldsymbol{z}. \tag{4.2.13}$$

From the Taylor's expansion in equation (4.2.8), and by the Maximum Likelihood Principle (MLP) that

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})\frac{\partial \ln P(\mathbf{y} \mid \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mid_{\boldsymbol{\beta}_{ML}} = 0,$$

it follows that

$$\ln P(\mathbf{y} \mid \boldsymbol{\beta}) = \ln P(\mathbf{y} \mid \boldsymbol{\beta}_{\mathbf{M}}) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})^{'}\mathbf{H}(\boldsymbol{\beta} - \boldsymbol{\beta}_{ML}) \mid_{\boldsymbol{\beta}_{ML}} \tag{4.2.14}$$

Taking exponent on both side of equation (4.2.14), we have

$$P(\mathbf{y} \mid \boldsymbol{\beta}) = P(\mathbf{y} \mid \boldsymbol{\beta}_{ML}) \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})^{'}\mathbf{H}(\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})\right] \tag{4.2.15}$$

or

$$P(\mathbf{y} \mid \boldsymbol{\beta}) = \ell_{ML} \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})^{'}\mathbf{H}(\boldsymbol{\beta} - \boldsymbol{\beta}_{ML})\right], \tag{4.2.16}$$

where

$$\ell_{ML} = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{z} - \mathbf{M}\boldsymbol{\beta}_{ML})^{'}(\boldsymbol{z} - \mathbf{M}\boldsymbol{\beta}_{ML})\right]. \tag{4.2.17}$$

4.2.4 Posterior Function Of Gaussian Process Regression

Assuming that the prior distribution of $\boldsymbol{\beta}$ is

$$p\left(\boldsymbol{\beta}\right) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{'}\boldsymbol{\Sigma}_{\beta}^{-1}\boldsymbol{\beta}\right). \tag{4.2.18}$$

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

$$P(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) \propto \exp\left[-\frac{1}{2}\left(\mathbf{y} - \mathbf{X}^{'}\boldsymbol{\beta}\right)^{'}\boldsymbol{\Sigma}^{-1}\left(\mathbf{y} - \mathbf{X}^{'}\boldsymbol{\beta}\right)\right] \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{'}\boldsymbol{\Sigma}_{\beta}^{-1}\boldsymbol{\beta}\right) \tag{4.2.19}$$

Simplifying equation (4.2.19), it follows that

$$P(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) \propto \exp\left[-\frac{1}{2}\left(\mathbf{y}^{'}\nu\mathbf{y} - 2\boldsymbol{\beta}^{'}\boldsymbol{X}^{'}\nu\mathbf{y} + \boldsymbol{\beta}^{'}\left(\boldsymbol{X}^{'}\nu\boldsymbol{X} + \kappa\mathbf{I}\right)\boldsymbol{\beta}\right)\right], \tag{4.2.20}$$

where $\nu = \boldsymbol{\Sigma}^{-1}$ and $\kappa = \boldsymbol{\Sigma}_{p}^{-1}$ are the covariance of the likelihood function and the prior function respectively.

The equation (4.2.20) above is indeed Gaussian with the constant term $\mathbf{y}^{'}\mathbf{y}$

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" [Gill, 2002]. The exponent of a general Gaussian distribution defined as $(\mathbf{y}-\mu)^{'}\mathbf{\Lambda}(\mathbf{y}-\mu)$, where $\mathbf{\Lambda}$ is the precision matrix can be expressed as

$$(\mathbf{y}-\mu)^{t}\mathbf{\Lambda}(\mathbf{y}-\mu)=\left[\mathbf{y}^{'}\mathbf{\Lambda}\mathbf{y}-2\mathbf{y}^{'}\mathbf{\Lambda}\mu+\mu^{'}\Lambda\mu\right]=\mathbf{y}^{'}\mathbf{\Lambda}\mathbf{y}-2\mathbf{y}^{'}\mathbf{\Lambda}\mu+\text{constant}. \tag{4.2.21}$$

Comparing equation (4.2.20) with equation (4.2.21) , we have

$$\mathbf{\Lambda}=\mathbf{X}^{'}\nu\mathbf{X}+\kappa\mathbf{I},\quad \mu=\nu\mathbf{\Lambda}^{-1}\mathbf{X}^{'}\mathbf{y} \tag{4.2.22}$$

That is $P(\boldsymbol{\beta}\mid\mathbf{y},\mathbf{X})\sim N(\boldsymbol{\beta}\mid\mu,\mathbf{\Lambda}^{-1})$. Note that $\mathbf{\Lambda}^{-1}$ must be invertible, that is $\mid\mathbf{\Lambda}\mid\neq 0$.

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

$$MAP=\nu\left(\nu\mathbf{X}^{'}\mathbf{X}+\kappa\mathbf{I}\right)^{-1}\mathbf{X}^{'}\mathbf{y}=\Sigma^{-1}\left(\Sigma^{-1}\mathbf{X}^{'}\mathbf{X}+\Sigma_{\beta}^{-1}\mathbf{I}\right)^{-1}\mathbf{X}^{'}\mathbf{y}. \tag{4.2.23}$$

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

$$\widehat{\boldsymbol{\beta}}=\Sigma^{-1}\left(\Sigma^{-1}\mathbf{X}^{'}\mathbf{X}+\Sigma_{\beta}^{-1}\mathbf{I}\right)^{-1}\mathbf{X}^{'}\mathbf{y}\quad\text{and}\quad\mathbf{\Lambda}^{-1}=\left(\Sigma^{-1}\mathbf{X}^{'}\mathbf{X}+\Sigma_{\beta}^{-1}\mathbf{I}\right)^{-1} \tag{4.2.24}$$

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

$$P(\boldsymbol{\beta}|\mathbf{y},\mathbf{X})\propto\exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^{'}\mathbf{\Lambda}^{-1}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\right). \tag{4.2.25}$$

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

$$P(\boldsymbol{\beta})=\left(\frac{1}{2\pi}\right)^{P/2}\left(\frac{1}{\tau_{\beta}}\right)^{P}\exp\left(-\frac{1}{2}\sum_{p=0}^{P}\frac{\beta_{p}^{2}}{\tau_{\beta}^{2}}\right) \tag{4.2.26}$$

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

$$P(\boldsymbol{u})=\left(\frac{1}{2\pi}\right)^{n/2}\left(\frac{1}{\tau_{u}}\right)^{n}\exp\left(-\sum_{i=1}^{n}\frac{u_{i}^{2}}{2\tau_{u}^{2}}\right). \tag{4.2.27}$$

Therefore, the posterior distribution is defined as

$$P\left(\boldsymbol{\beta},\boldsymbol{u},\tau_{\beta}^{2},\tau_{u}^{2}\mid\mathbf{y},\mathbf{E},\vartheta\right)\propto P(\mathbf{y},\mathbf{E}\mid\boldsymbol{\vartheta},\boldsymbol{\beta},\boldsymbol{u})\,P(\boldsymbol{\beta})\,P(\boldsymbol{u}) \tag{4.2.28}$$

Hence,

$$P\left(\beta, \mathbf{u}, \tau_\beta^2, \tau_u^2 \mid \mathbf{X}, \mathbf{E}, \vartheta\right) = \prod_{i=1}^{n} \frac{\left(E_i \exp(\eta_i)\right)^{y_i} \exp\left(-E_i \exp(\eta_i)\right)}{y_i!} \times \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta^2}\right)^{P} \exp\left(-\frac{1}{2}\sum_{p=0}^{P}\frac{\beta_p^2}{\tau_\beta^2}\right) \times \left(\frac{\tau_u^2}{2\pi}\right)^{1/2} \exp\left(-\sum_{i=1}^{n}\frac{u_i^2}{2\tau_u^2}\right). \tag{4.2.29}$$

The prior distribution for the linear regression coefficients are is given by $\boldsymbol{\beta} \sim N\left(0, \tau_\beta^2\right)$. The corresponding conjugate prior distribution for τ_β^2 is the inverse-gamma defined by [Gill, 2002, Ntzoufras, 2011]

$$P\left(\tau_\beta^2 \mid \omega, \phi\right) = \frac{\phi^\omega}{\Gamma(\omega)} \left(\tau_\beta^2\right)^{-(\omega+1)} \exp\left(-\frac{\phi}{\tau_\beta^2}\right), \tau_\beta^2, \omega, \phi > 0. \tag{4.2.30}$$

The equation (4.2.30) is the hyperprior distribution for τ_β^2 with hyperparameters (ω, ϕ). We defined $\tau_\beta^2 \sim Gamma\left(0.05, 0.005\right)$ and modelled the random effect $u_i \sim N\left(0, \tau_u^2\right)$ and hyperprior distribution for the precision parameter $\tau_u^2 \sim Gamma\left(0.05, 0.005\right)$.

Parameter estimation was carried out using Bayesin Markov Chain Monte Carlo via Gibbs Sampling (See Section 3.4). MCMC convergence was reached at 100,000 iterations after a burnin period of 10,000 sample and thinning of every 30^{th} element in the sample. Figure 4.5 presents convergence diagnostics plots of this model.

4.2.5 Markov Chain Monte Carlo Diagnosis

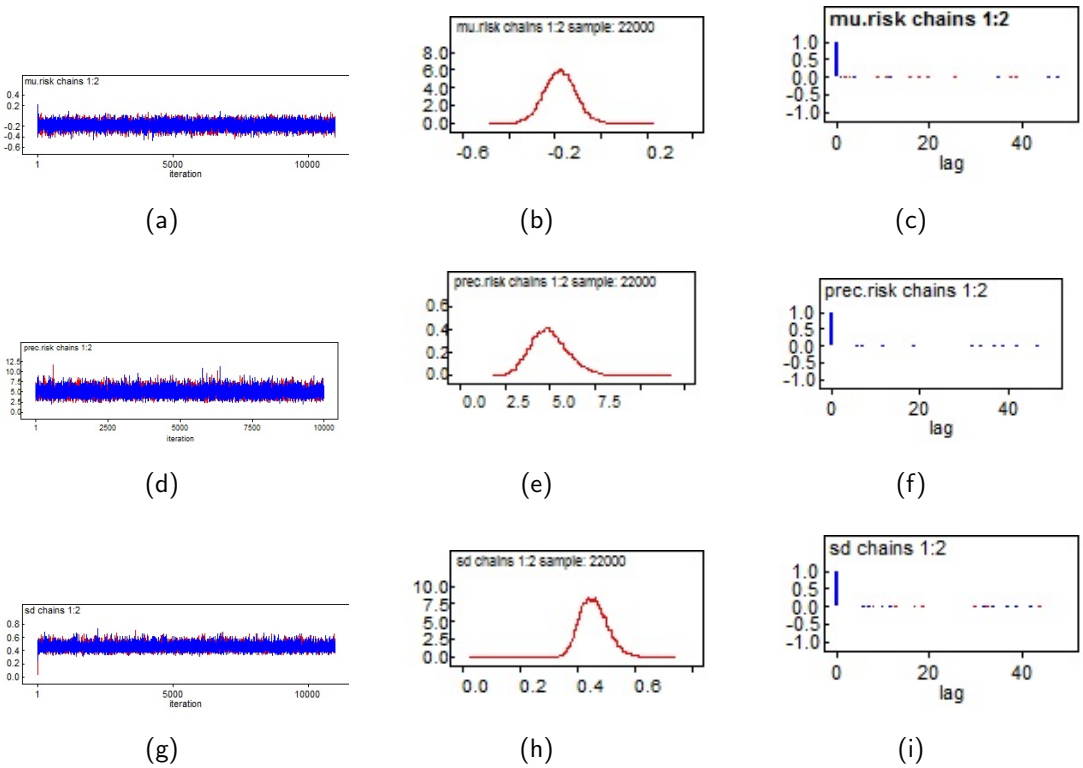


Figure 4.5: Poisson Log-Normal Model: Convergence Diagnosis of Markov Chain Monte Carlo

Figure 4.5 presents Gelman and Rubin convergence diagnostics of the Poisson Log-Normal model without covariate and random effect: Column-wise from the top left, Figure 4.5 (a)-(g) are trace plots for the mean, precision, and the standard deviation respectively. Figure 4.5 (b)-(h) are posterior marginal density plots for the mean, precision, and the standard deviation respectively. Figure 4.5 (c)-(i) are autocorrelation plots for the mean, precision, and the standard deviation 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. Hence, the density plots for the parameters show that convergence of the chain has reached.

Posterior statistics of this model fitted to Kenya TB data is presented in Table 4.2.

4.2.6 Results of the Poisson Log-Normal Model Without Covariate and Random Effect

This section presents the results for fitting the Poisson Log-Normal model to Kenya TB data. Table 4.2 presents a summary of the Poisson Log-Normal model fitted.

Table 4.2: Posterior Statistics of the Poisson Log-Normal Model.

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	-

Table 4.2, 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 [Lawson et al., 2003].

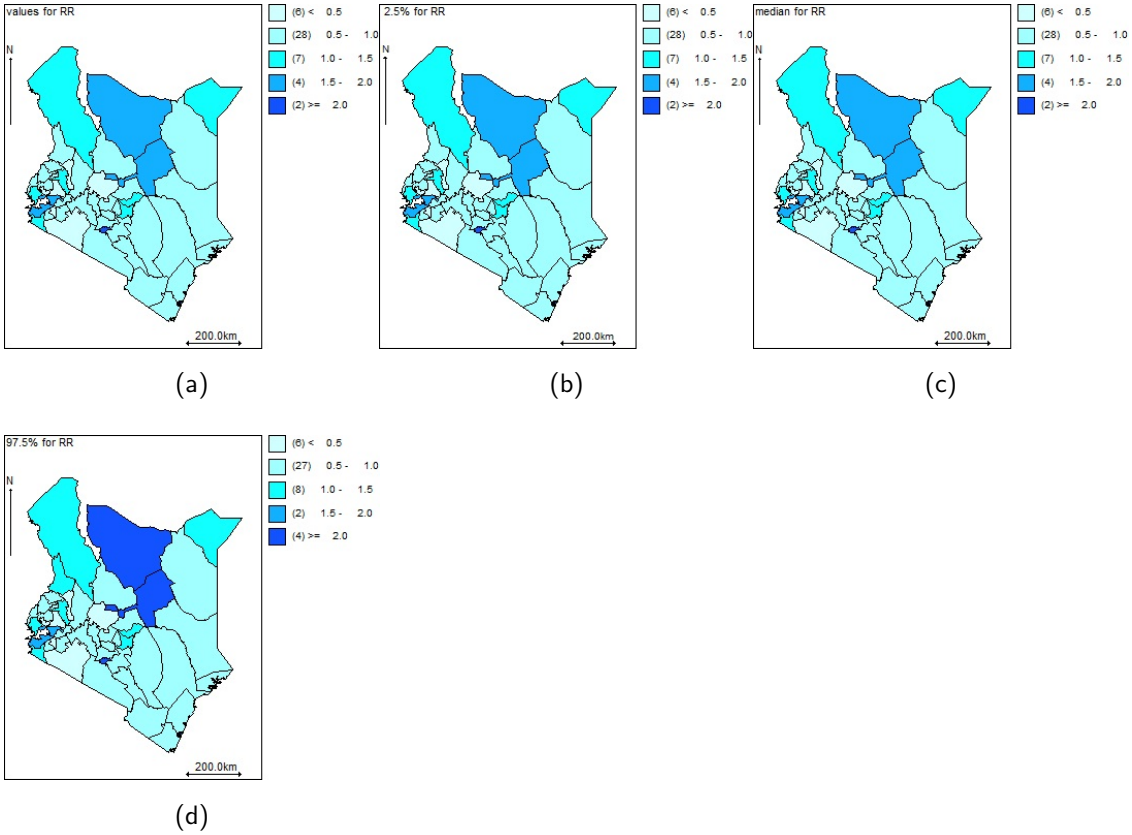


Figure 4.6: Kenya County Level TB Prevalence Counts: Poisson Log-Normal model Without covariate and random effect. (4.6a) The posterior relative risk map and its 2.5% quantile (4.6b), median (4.6c) and 97.5% quantile (4.6d).

Figure 4.6, revealed that TB risk is expected to be high in the North,West, North-West and Central counties of 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: 4.7 below. 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.

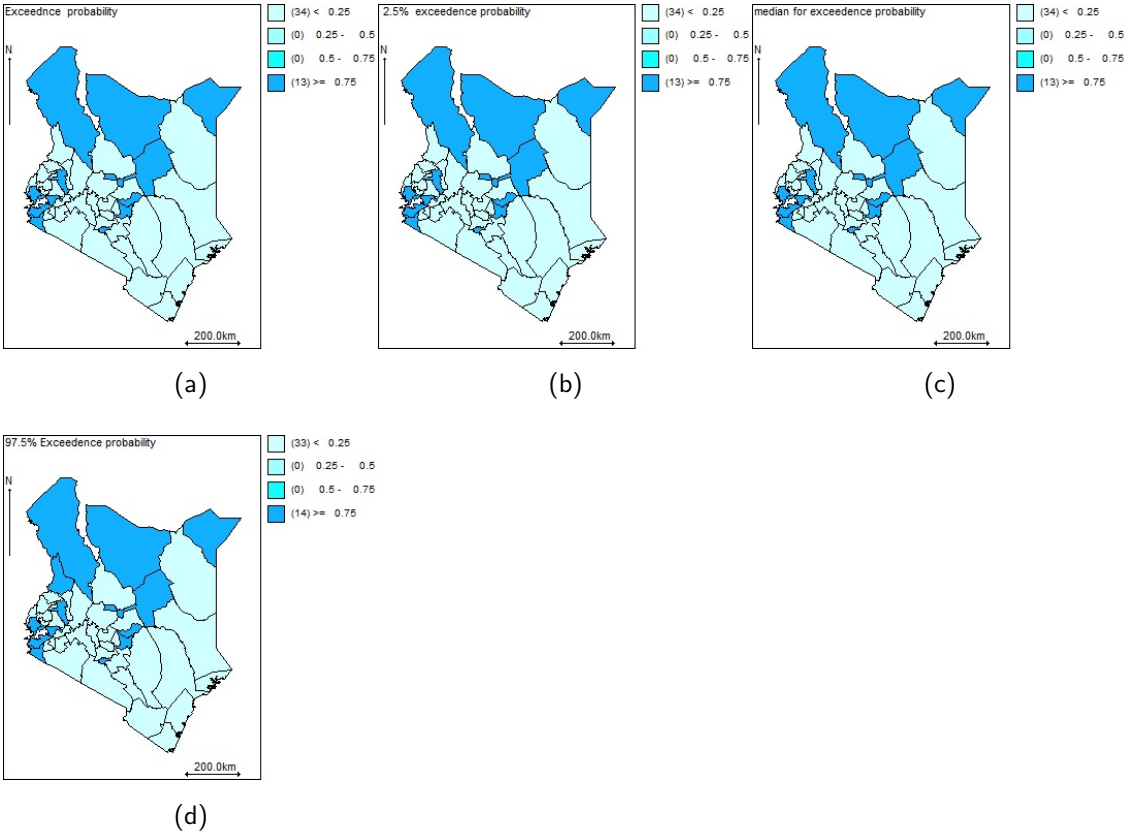


Figure 4.7: Kenya County Level TB Prevalence Counts: Poisson Log-Normal model Without covariate and random effect. (4.7a) The posterior relative risk exceedence probability map and its 2.5% quantile (4.7b), median (4.7c) and 97.5% quantile (4.7d).

The exceedence probability map Figure 4.7 of the Poisson Log-Normal model also 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 4.7 confirms with Figure 4.6 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.

We also present the MCMC convergence diagnostics test of the UH model in the next section.

4.2.7 Markov Chain Monte Carlo Diagnostics

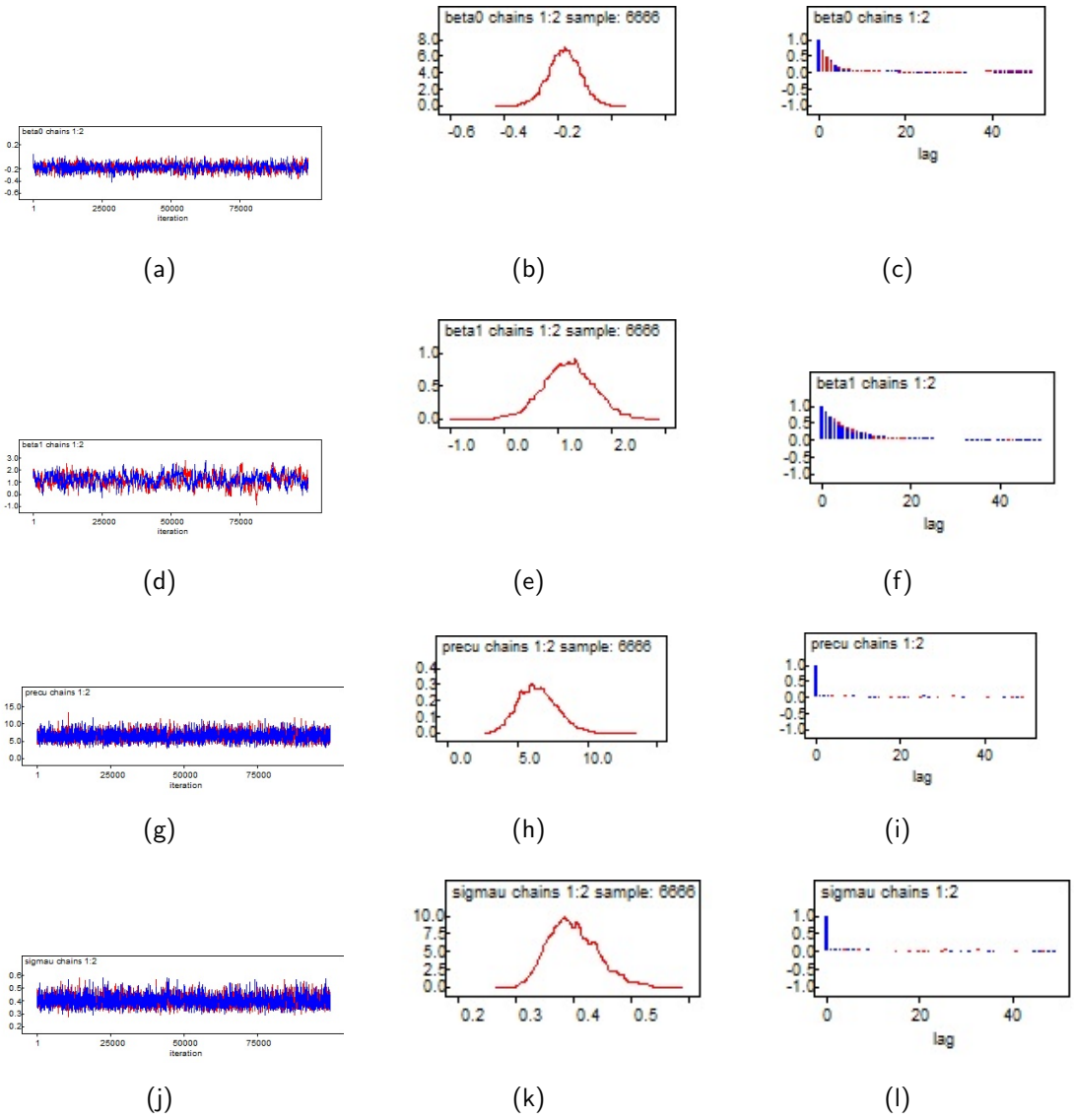


Figure 4.8: Poisson Log-Normal Model with Uncorrelated Heterogeneity (UH) Effect: Convergence Diagnosis of Markov Chain Monte Carlo

Figure 4.8 presents Gelman and Rubin convergence diagnostics of the Poisson Log-Normal model without covariate and random effect: Column-wise from the top left, Figure 4.8 (a)-(j) are trace plots for the intercept, the HIV covariate parameter, the precision, and the standard deviation respectively. Figure 4.8 (b)-(k) are posterior marginal density plots for the intercept, the HIV covariate parameter, the precision, and the standard deviation respectively. Figure 4.8 (c)-(l) are

autocorrelation plots for for the intercept, the HIV covariate parameter, the precision, and the standard deviation 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. Hence, the density plots for the parameters show that convergence of the chain has reached.

4.2.8 Application of the Poisson-Log-Normal with UH Effect and Covariates

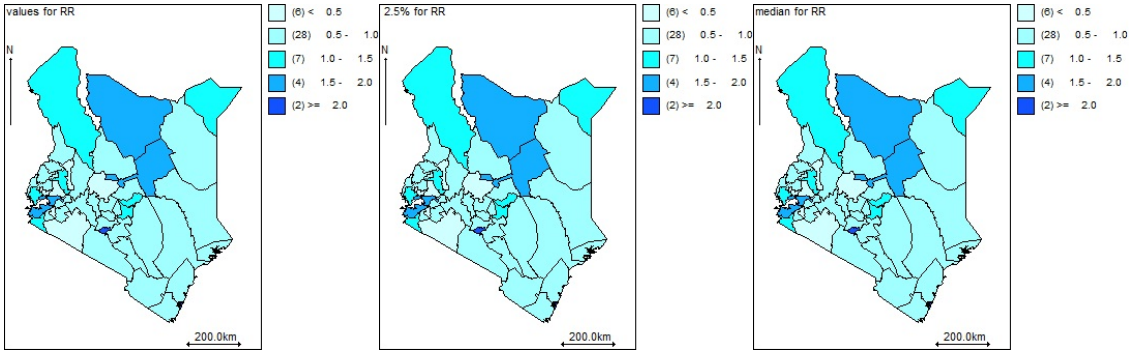
This section presents the results for fitting the Poisson Log-Normal model with county level UH effects to Kenya TB data.

Table 4.3: Posterior Statistics of the Poisson Log-Normal Model with UH Effect.

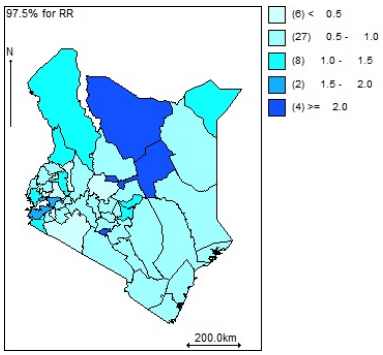
Model indicators	UH
β_0	-0.1765(-0.2957,-0.05936)
HIV	1.198 (0.4928,2.571)
Firewood	0.2735(-2.215,2.144)
five kilometer distance	-1.317(-3.423,1.437)
σ_u	0.4409(0.359,0.5466)
τ_u^2	5.324(3.347,7.757)
$\bar{D}$	575,779
pD	46.973
DIC	622.753

Table 4.3 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 covariates 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. Lawson et al. [2003] 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 τ_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 τ^2 = 5.01 (95% credible interval = 3.18-7.24).



(a) (b) (c)



(d)

Figure 4.9: Kenya County Level TB Prevalence Counts: UH smooth relative risk map (4.9a) and its 2.5% quantile (4.9b), median (4.9c) and 97.5% quantile (4.9d).

The Figure 4.9 also confirmed that out of the 47 counties in Kenya, 13 exhibit TB relative risk higher than the national risk (RR=1). Table 4.4 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 4.4: 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)	-

Table 4.4 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 4.10 below visually presents counties with risk above 1.

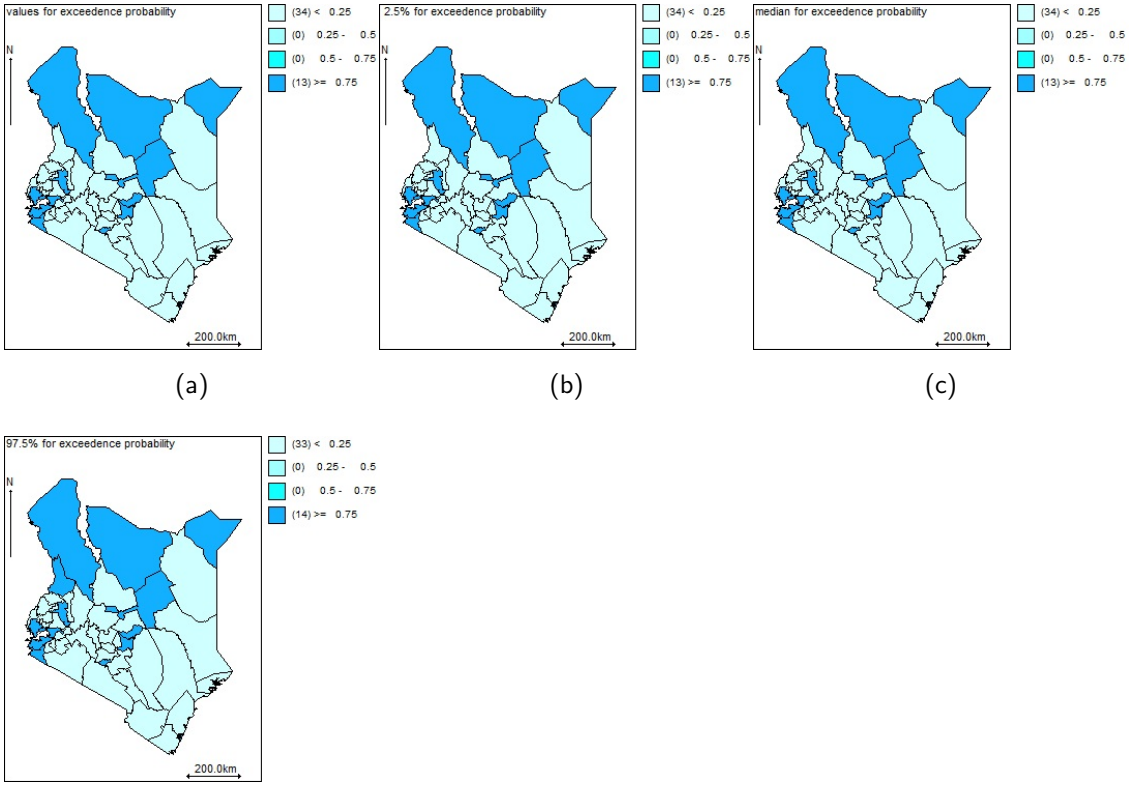


Figure 4.10: Kenya County Level TB Prevalence Counts: UH smooth relative risk exceedence probability map (4.10a) and its 2.5% quantile (4.10b), median (4.10c) and 97.5% quantile (4.10d).

The Figure 4.10 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.

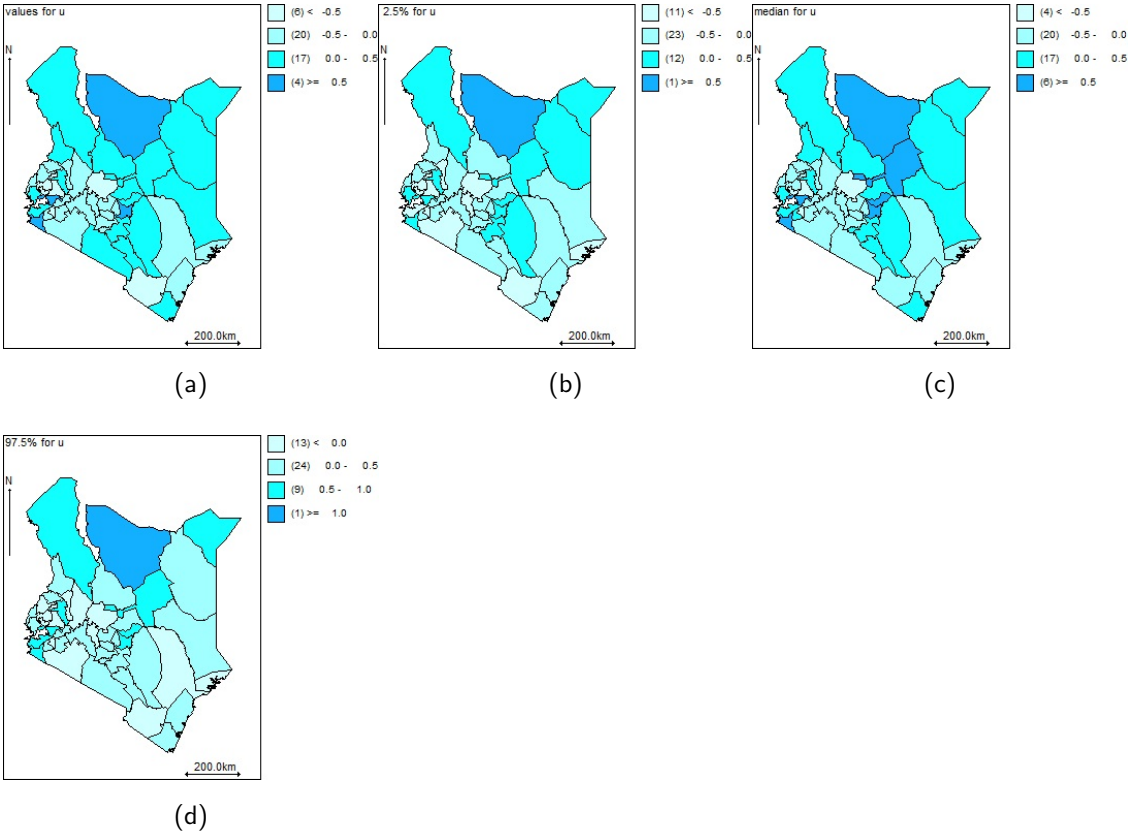


Figure 4.11: Area-Specific Random Effect: The posterior map (4.11a) and its 2.5% quantile (4.11b), median (4.11c) and 97.5% quantile (4.11d).

Figure 4.11 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 [Lawson et al., 2003]. Excess risk of TB is observed in Marsbit, Embu, Migori and Kisumu.

4.3 Summary of the Nonspatial 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 [Lawson et al., 2003]. The Poisson Log-Normal model provides specifications that allow for incorporation of covariates. The Poisson Log-Normal model also enable us to cap-

ture 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 [Gill, 2002]. Gill [2002] 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 lower 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 [Gill, 2002]. Nevertheless, considerable work on convergence diagnostics has been done to make specific recommendations and identify tests [Gill, 2002].

HIV is identified as significant among the covariates considered. Reason being that HIV patients have their immune system weakened or destroyed by the HIV virus 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 5.

Chapter 5

Bayesian Hierarchical Spatial Model for Use in Disease Mapping.

This chapter 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 autocorrelation 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*" Miller [2004].

In particular, we investigate the statistical properties of the Conditional autoregressive (CAR) model and the Julian Besag [1991] models.

5.1 Conditional Autoregressive (CAR) Model

Though conditional autoregressive models where introduced decades ago by Julian Besag [1991], they were not widely used until the 1990s. Since they are defined conditionally, they are particularly suited for use with the Gibbs sampler [Geman and Geman, 1984].

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 [Lawson et al., 2003]. The distances or boundaries between the regions are often used in determine neighbourhood properties within CAR models [Kyung and Ghosh, 2009].

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 [Besag, 1974, Cressie, 1993]. Application of CAR models can found in [Besag, 1974, Smith and R. L, 2001, Mariella and Tarantino, 2010].

Let $S = \{1, 2, \ldots, n\}$ represents the 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 $\boldsymbol{v}$ as the vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)'$.

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

$$v_i \mid v_{j \neq i} \sim N\left(\sum_{i \neq j} \Phi_{ij} v_j, \tau_i^2\right). \tag{5.1.1}$$

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 $\Phi_{ij} = \frac{1}{M}$ 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 [Mariella and Tarantino, 2010].

The Gaussian processes are specified by their mean and covariance function [Waller and Gotway, 2004]. 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

$$E\left[v_i \mid v_{j \neq i}\right] = \mu_i + \sum_{j \in N_i} \Phi_{ij}\left[v_j - \mu_i\right] \quad \text{and} \quad var\left(v_i \mid v_j\right) = \tau_i^2. \tag{5.1.2}$$

Therefore, conditional probability density function of a CAR random variable v_i is has the form [Mariella and Tarantino, 2010]

$$f\left(v_i \mid v_{j\neq i} \in S\right) = \sqrt{\frac{1}{2\pi\tau_i^2}} \exp\left\{ -\frac{\left[\left(v_i - \mu_i\right) - \rho \sum\limits_{j\in N_i} \Phi_{ij}\left(v_j - \mu_j\right)\right]^2}{2\tau_i^2} \right\}, \tag{5.1.3}$$

where $\mu_i = \in \mathbb{R}, \tau_i^2 \in \mathbb{R}^+, \mid \rho \mid < 1,\ \Phi_{(ij)} \in \mathbb{R}, \Phi_{(ij)} = \Phi_{(ji)},\ \Phi_{(ii)} = 0.$

The conditional joint probability density function of all the observations is

$$f\left(v_i \mid v_{j\neq i}\right) = \frac{1}{(2\pi)^{n/2}\, det\left(\mathbf{B^{-1}\Sigma_D}\right)^{1/2}} \exp\left[-\frac{\left(\boldsymbol{v}-\boldsymbol{\mu}\right)' \mathbf{\Sigma_D^{-1}B}\left(\boldsymbol{v}-\boldsymbol{\mu}\right)}{2} \right], \tag{5.1.4}$$

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

$$\mathbf{B} = (\mathbf{I} - \rho\mathbf{\Phi}) \quad \text{with} \quad B_{(ij)} = \begin{cases} 1 & \text{if} \quad i = j \\ -\rho\Phi_{(ij)} & \text{if} \quad j \in N_i \\ 0 & \text{otherwise} \end{cases}, \tag{5.1.5}$$

$\mathbf{\Sigma_D} \in \mathbb{R}^{+n\times n}$ diagonal matrix; $\mathbf{\Sigma_D} = \quad diag\left(\tau_1^2, \ldots, \tau_n^2\right) = \tau_i^2$ such that $\mathbf{\Sigma_D}$ is symmetric. It follows that the joint multivariate Gaussian distribution for v_i with $\mu = 0$ has covariance matrix $\mathbf{\Sigma_{D_v}} = \mathbf{\Sigma_D^{-1}B} = \mathbf{B^{-1}\Sigma_D}$ which is symmetric such that $\Phi_{(ij)}\tau_j^2 = \Phi_{(ji)}\tau_i^2, i, j \in S$.

Thus, a conditional autoregressive model $\boldsymbol{v}$ in (5.1.3) has a probability density function defined as

$$v_i \mid v_{j\neq i} \sim N\left[\mu_i + \rho \sum_{j\in N_i} \Phi_{(ij)}\left(v_j - \mu_j\right), \tau_i^2\right], i \in S \tag{5.1.6}$$

and the joint probability density function in (5.1.4) becomes

$$\boldsymbol{v} \sim N\left(\mu, \mathbf{B^{-1}\Sigma_D}\right). \tag{5.1.7}$$

The necessary and sufficient condition for (5.1.7) 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 $\lambda_i > 0, i, \ldots, n)$ [Mariella and Tarantino, 2010]. For v_i to be a Gaussian random variable, we need to show that $\mathbf{\Sigma_D}$ is symmetric.

To show that $\mathbf{\Sigma_D}$ is symmetric, we defined a symmetric weighted adjacency matrix $\mathbf{W}$, where

$$\mathbf{W} = \left(w_{(ij)}\right) \quad \text{with} \quad w_{(ij)} = \begin{cases} 1 \quad \text{if} \quad j = i \\ \varphi\left(i,j\right) \quad \text{with} \quad j \in N_i : \forall i,j \in S, w_{(ij)} = w_{(ji)} \\ 0 \quad \text{otherwise,} \end{cases} \tag{5.1.8}$$

where $\varphi\left(i,j\right)$ is a measure that quantifies the proximity between region i and region j; if $\varphi\left(i,j\right) = 1$, then i and j share a common boundary (neighbours). The $\varphi\left(i,j\right)$ could be the distance between the centroids of region i and j. Also, if $\varphi\left(i,j\right) = 1$, then j is one of the h nearest neighbours of i. Let $\mathbf{W_D}$ be the diagonal of the adjacency matrix $\mathbf{W}$. The adjacency matrix of normalization or standardization $\mathbf{W_D}$ is defined as

$$\mathbf{W_D} = \text{diag}\left(w_{(1+)}, w_{(2+)}, \ldots, w_{(n+)}\right). \tag{5.1.9}$$

Suppose

$$w_{(i+)} = \sum_{j \in N_i} w_{(ij)}, i,j \in S, \tag{5.1.10}$$

then we define a matrix of interaction, $\mathbf{\Theta}$ to be a normalized adjacency matrix defined as

$$\mathbf{\Theta} = \mathbf{W_D^{-1}W} \quad \text{with} \quad \Theta_{(ij)} = \frac{w_{(ij)}}{w_{(i+)}}, \quad \text{where} \quad \Theta_{(ij)}\tau_j^2 = \Theta_{(ji)}\tau_i^2, i,j \in S. \tag{5.1.11}$$

Suppose again that the matrix $\mathbf{W_D}$ corresponding to a constant diagonal matrix normalized as (5.1.11), then we have

$$\mathbf{W_D} = \tau^2\mathbf{W_D^{-1}} \quad \text{with} \quad \tau_i^2 = \frac{\tau^2}{w_{(i+)}}, i \in S, \tau^2 \in \mathbb{R}^+. \tag{5.1.12}$$

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

$$P\left(v_1, \ldots, v_n\right) \propto \exp\{-\frac{1}{2\tau^2}\mathbf{\Psi}'\left(\mathbf{\Sigma_{D_w}} - \mathbf{W}\right)^{-1}\mathbf{\Sigma_D\Psi}\}, \tag{5.1.13}$$

where $\mathbf{\Psi} = (\boldsymbol{v} - \boldsymbol{\mu})$ and $\mathbf{B} = (\mathbf{\Sigma_{D_w}} - \mathbf{W})$.

Hence, the CAR model structure for v_i is defined as

$$v_i \mid v_{j\neq i} \sim N\left(\sum_j \frac{w_{ij}}{w_{i+}} v_j, \frac{\tau_v^2}{w_{i+}}\right), \tag{5.1.14}$$

and the equa tion (5.1.7) can be alternatively defined as

$$\boldsymbol{v} \sim \left(\boldsymbol{\mu}, \left[\frac{1}{{\tau_v}^2}\left(\mathbf{W_D} - \rho\mathbf{W}\right)\right]^{-1}\right). \tag{5.1.15}$$

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 [Mariella and Tarantino, 2010].

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

$$\mathbf{\Sigma_{D_v}^{-1}} = \mathbf{\Sigma_{D_w}}\left(\mathbf{I} - \rho\widehat{\mathbf{W}}\right). \tag{5.1.16}$$

If $\mid \rho \mid < 1$, then $\mathbf{\Sigma_{D_w}}\left(\mathbf{I} - \rho\widehat{\mathbf{W}}\right)$ 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 $\mathbf{\Sigma_{D_w}^{-1/2} W \Sigma_{D_w}^{-1/2}}$. 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 [Banerjee, 2009].

5.1.1 Parameter Estimation

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

$$y_i \sim Poisson\left(E_i \exp\left(\eta_i\right)\right), \quad \text{were} \quad \mu_i = E_i \exp\left(\eta_i\right) \quad \text{is the mean of Poisson distribution.} \tag{5.1.17}$$

The relative risk is defined as

$$\vartheta_i = \exp(\eta_i), \quad \text{where} \quad \eta_i = \mathbf{X}'\boldsymbol{\beta} + v_i, \quad \text{and} \quad v_i \quad \text{has a CAR structure.} \tag{5.1.18}$$

Fitting the generalized linear model with a log-link function, we have $\log(\mu_i) = \log(E_i) + \boldsymbol{X}'\boldsymbol{\beta} + v_i$

Under the Bayesian method, given the likelihood function of $\boldsymbol{\vartheta}$ defined as

$$\ell(\beta, \mathbf{v}) = \prod_{i=1}^{n} \frac{(E_i \exp(\eta_i))^{y_i} \exp(-E_i \exp(\eta_i))}{y_i!} = P(\mathbf{y}, \mathbf{E}, \vartheta \mid \boldsymbol{\beta}, \boldsymbol{v}), \tag{5.1.19}$$

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

$$P(\boldsymbol{\beta}) = \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^{P} \exp\left(-\frac{1}{2}\sum_{p=0}^{P}\frac{\beta_p^2}{\tau_\beta^2}\right) \tag{5.1.20}$$

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

$$P(\boldsymbol{v}) = \left[v_i \mid v_{j\neq i}, \tau_v^2\right] \sim N\left(\sum_{j\neq i}\frac{w_{ij}}{w_{ij}} v_j, \frac{\tau_v^2}{w_{ij}}\right) \sim CAR\left(0, \tau_v^2\right). \tag{5.1.21}$$

The posterior distribution is defined as

$$P\left(\boldsymbol{\beta}, \boldsymbol{v}, \tau_\beta^2, \tau_v^2 \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta}\right) \propto P\left(\mathbf{X}, \mathbf{E}, \boldsymbol{\vartheta} \mid \boldsymbol{\beta}, \boldsymbol{v}, \tau_\beta^2, \tau_v^2\right) P(\boldsymbol{\beta}) P(\boldsymbol{v}). \tag{5.1.22}$$

Therefore,

$$\begin{aligned} P\left(\boldsymbol{\beta}, \boldsymbol{v}, \tau_\beta^2, \tau_v^2 \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta}\right) = \prod_{i=1}^{n} \frac{(E_i \exp(\eta_i))^{y_i} \exp(-E_i \exp(\eta_i))}{y_i!} \times \\ \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^{P} \exp\left(-\frac{1}{2}\sum_{p=0}^{P}\frac{\beta_p^2}{\tau_\beta^2}\right) \times \\ \left(\sum_{j\neq i}\frac{w_{ij}}{w_{ij}} v_j, \frac{\tau_v^2}{w_{ij}}\right). \end{aligned} \tag{5.1.23}$$

The hyperperprior distribution for the precision parameters τ_v^2 and τ_β^2 are $\tau_v^2 \sim Gamma(0.05, 0.005)$ and $\tau_\beta^2 \sim Ggamma(0.5, 0.05)$ 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 (See Section 3.4). Convergence of the MCMC was reached at 11000 iteration after a burnin period of 10,000 sample and thinning of every 30^{th} element of the chain. Convergence diagnosis plots are presented in Figure 5.1 and posterior statistics of parameters presented in Table 5.1.

5.2 The Besag, York and Mollié (BYM) Model

Among the models proposed for performing risk smoothing which have appeared in literature, the Julian Besag [1991] 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 (as already discussed in Section 5.1 and Section 4.2 respectively).

BYM model was introduced by Clayton and Kaldor [1987] and latter developed by Julian Besag [1991]. The BYM or convolution model is defines as

$$\eta_i = \mu + u_i + v_i \tag{5.2.1}$$

As noted earlier, assume

$$y_i \sim Poisson\left(E_i \exp\left(\eta_i\right)\right), \quad \text{where} \quad \mu_i = E_i \exp\left(\eta_i\right) \quad \text{is the mean of the Poisson distribution.} \tag{5.2.2}$$

The linear link function $\eta_i = \mathbf{X}'\boldsymbol{\beta} + u_i + v_i$. The log relative risk $\log\left(\vartheta_i\right) = \eta_i$. Therefore, the relative risk for the i area is defined by

$$\vartheta_i = \exp(\mathbf{X}'\boldsymbol{\beta} + u_i + v_i). \tag{5.2.3}$$

Finding a generalized linear model with-link function, we have

$$\log\left(\mu_i\right) = \log\left(E_i\right) + \exp\left(\mathbf{X}'\boldsymbol{\beta} + u_i + v_i\right) \tag{5.2.4}$$

Where $\mathbf{y}, \boldsymbol{\beta}, \mathbf{E}$ and $\boldsymbol{\vartheta}$ 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.

5.2.1 Parameter Estimation

Defined the likelihood function as

$$\ell\left(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}\right) = \prod_{i=1}^{n} \frac{\left(E_i \exp(\eta_i)\right)^{y_i} \exp\left(-E_i \exp(\eta_i)\right)}{y_i!} = P\left(\mathbf{y}, \mathbf{E}, \vartheta \mid \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}\right). \tag{5.2.5}$$

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

$$P(\boldsymbol{\beta}) = \left(\frac{1}{2\pi}\right)^{P/2} \left(\frac{1}{\tau_\beta}\right)^{P} \exp\left(-\frac{1}{2}\sum_{p=0}^{P}\frac{\beta_p^2}{\tau_\beta^2}\right), \tag{5.2.6}$$

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

$$P(\boldsymbol{u}) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\tau_u}\right)^{n} \exp\left(-\sum_{i=1}^{n}\frac{u_i^2}{2\tau_u^2}\right), \tag{5.2.7}$$

and prior distribution for the CAR structure $v_i$is

$$P(\boldsymbol{v}) = \left[v_i \mid v_{j\neq i}, \tau_v^2\right] \sim N\left(\sum_{j\neq i}\frac{w_{ij}}{w_{ij}}v_j, \frac{\tau_v^2}{w_{ij}}\right) \sim CAR\left(0, \tau_v^2\right). \tag{5.2.8}$$

The posterior distribution is defined as

$$P\left(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \tau_\beta^2, \tau_u^2, \tau_v^2 \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta}\right) \propto P\left(\mathbf{y}, \mathbf{E}, \vartheta \mid \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \tau_\beta^2, \tau_u^2, \tau_v^2\right) P(\boldsymbol{\beta}) P(\boldsymbol{u}) P(\boldsymbol{v}). \tag{5.2.9}$$

Therefore,

$$\begin{aligned} P\left(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}, \tau_\beta^2, \tau_u^2, \tau_v^2 \mid \mathbf{y}, \mathbf{E}, \vartheta\right) = \prod_{i=1}^{n}\frac{(E_i\exp(\eta_i))^{y_i}\exp\left(-E_i\exp(\eta_i)\right)}{y_i!} \times \\ \left(\frac{1}{2\pi}\right)^{P/2}\left(\frac{1}{\tau_\beta}\right)^{P}\exp\left(-\frac{1}{2}\sum_{p=0}^{P}\frac{\beta_p^2}{\tau_\beta^2}\right) \times \\ \left(\frac{1}{2\pi}\right)^{n/2}\left(\frac{1}{\tau_u}\right)^{n}\exp\left(-\sum_{i=1}^{n}\frac{u_i^2}{2\tau_u^2}\right) \times \\ \left(\sum_{j\neq i}\frac{v_j w_{ij}}{w_{ij}}, \frac{\tau_v^2}{w_{ij}}\right). \end{aligned} \tag{5.2.10}$$

The hyperprior disribution for the precision parameters τ_u^2, τ_v^2 and τ_β^2 are $\tau_u^2 \sim Gamma\,(0.5, 0.005)$, $\tau_v^2 \sim Gamma\,(0.5, 0.005)$ and $\tau_\beta^2 \sim Gamma\,(0.5, 0.01)$ respectively. The linear regression coefficient are assumed to have normal distribution defined by $\boldsymbol{\beta} \sim N\left(0, \tau_\beta^2\right)$. The τ_u^2 reflects the amount of extra-poisson variation in the data [Lawson et al., 2003]. 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 (See Section 3.4). Convergence of the MCMC was reached at 11000 iteration after a burn-in period of 10,000 sample and thinning of every 90^{th} element of the chain. Posterior statistics of the CAR and the BYM model are presented in Table 5.1.

5.3 Application of CAR and BYM Models to the TB Data

This section presents the results for fitting the CAR model and the BYM model to Kenya TB prevalence data.

Table 5.1: 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

Table 5.1 revealed that the estimated overall relative risk effect of the CAR model is β_0 = -0.177(95% credible interval = (-0.180,-0.174)) and BYM model β_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 = 2.18-40.80). 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, but its robustness as a spatial model is lost a in situation where there is over-fitting [Aurélien et al., 2007]. 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 in the next section and its convergence diagnostics in Figure 5.1. Convergence diagnostics of the BYM model can be found in Appendix 7.

5.3.1 Markov Chain Monte Carlo Diagnostics

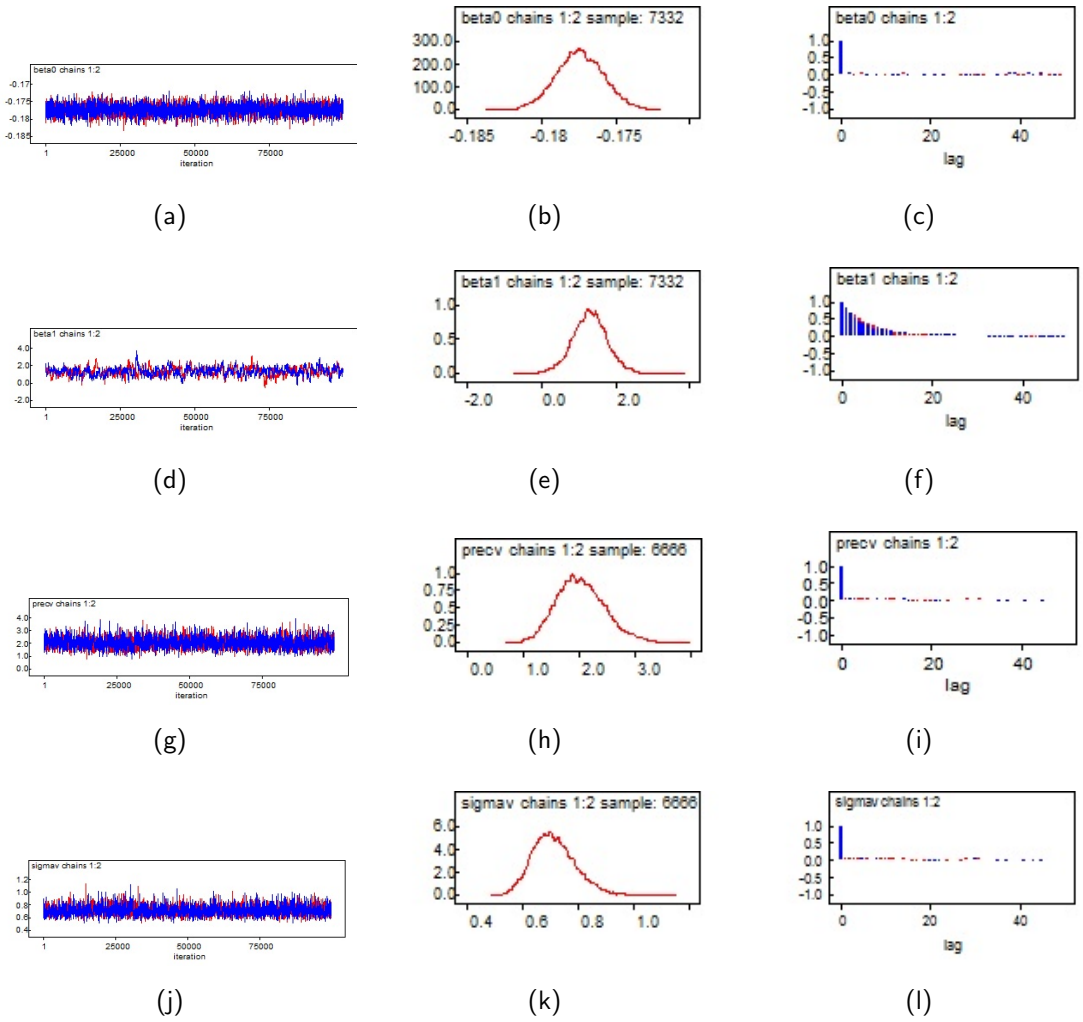


Figure 5.1: Poisson Log-Normal Model with CAR Model: Convergence Diagnosis of Markov Chain Monte Carlo

Figure 5.1 presents Gelman and Rubin convergence diagnostics of the CAR model: Column-wise from the top left, Figure 5.1 (a)-(j) are trace plots for the intercept, the HIV covariate parameter, the precision, and the standard deviation respectively. Figure 5.1 (b)-(k) are posterior marginal density plots for the intercept, the HIV covariate parameter, the precision, and the standard deviation respectively. Figure 5.1 (c)-(l) are autocorrelation plots for for the intercept, the HIV covariate parameter, the precision, and the standard deviation 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. Hence, the density plots for the parameters show that convergence of the chain has reached.

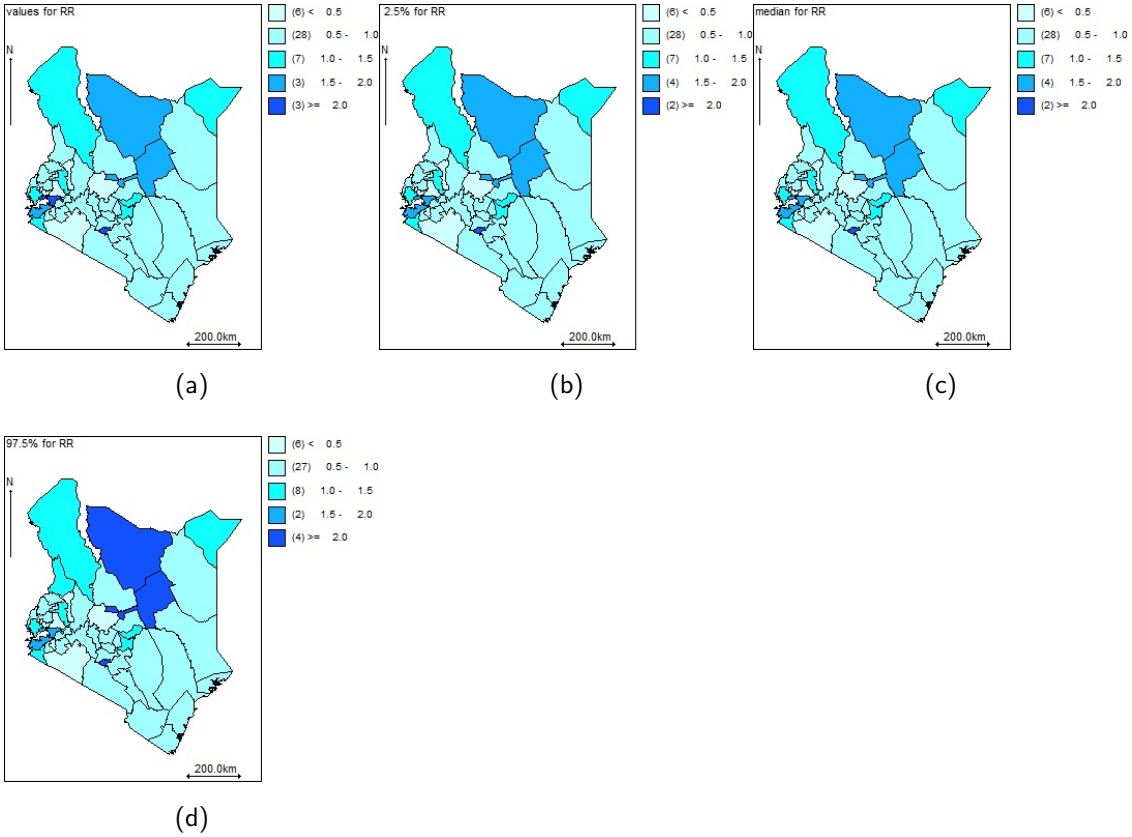


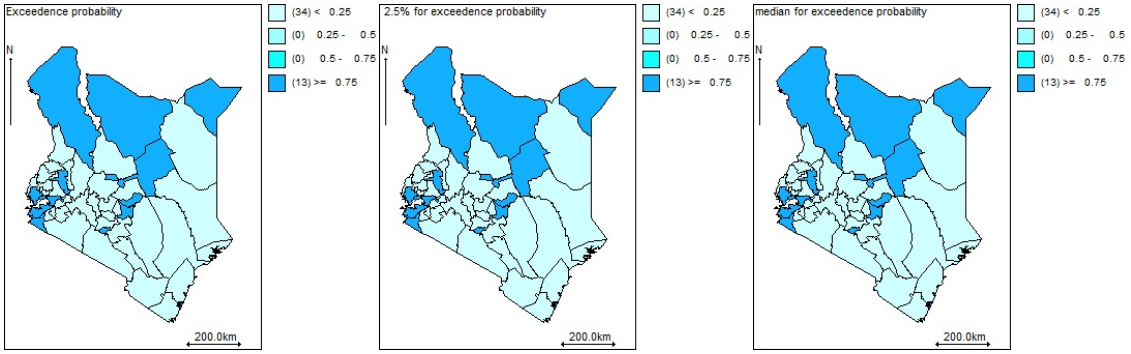
Figure 5.2: Kenya county level TB prevalence counts: The CAR model's posterior mean of the relative risk map (5.2a) and its 2.5% quantile (5.2b), median (5.2c) and 97.5% quantile (5.2d).

The Figure 5.2 also shows that out of the 47 counties in Kenya, 13 exhibit TB relative risk higher than the national risk (RR=1). Table 5.2 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 5.2: 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)	-

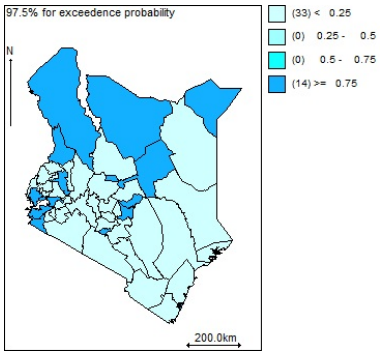
Table 5.2 shows 4 counties (Nairobi, Mombasa, Isiolo, and Marsabit) having highest relative risk (RR > 2.0). Out of the 47 counties, 13 counties show high relative risk above 1. Counties with high TB relative risk are visually shown by the exceedence probability map Figure 5.3 below.



(a)

(b)

(c)

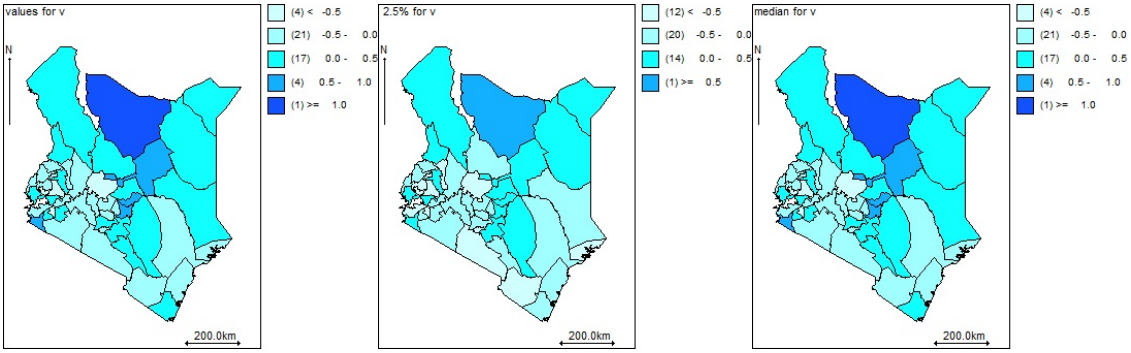


(d)

Figure 5.3: Kenya county level TB prevalence counts: The CAR model posterior mean of the relative risk exceedence probability map (5.3a) and its 2.5% quantile (5.3b), median (5.3c) and 97.5% quantile (5.3d).

Figure 5.3 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.

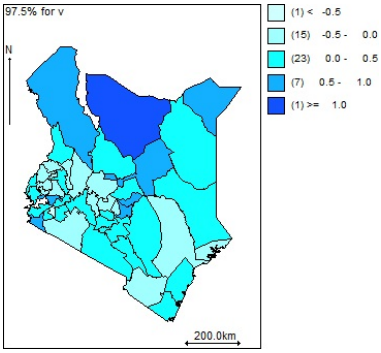
Again, it can be observed that North, West, North-West and Central counties of Kenya exhibit high TB prevalence and low prevalence in the South-West counties except Mombasa.



(a)

(b)

(c)



(d)

Figure 5.4: The Correlated Heterogeneity effect's posterior map (5.4a) and its 2.5% quantile (5.4b), median (5.4c) and 97.5% quantile (5.4d).

The Figure 5.4 is used to capture areas with potential clusters of disease risk. Clusters of TB risk is suspected in Marsabit, Embu, Migori and Kisumu.

5.4 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 when fitted with Kenya's TB data, but CAR model seems 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).

Chapter 6

Spatio-Temporal Modelling

This chapter presents models used for modelling risk in space and time. Many disease mapping models are restricted to identification of spatial heterogeneity and clustering of diseases risk which are in fact constrained to a single time period. However, most data in public health are often in the form of time window for several years. Therefore, there is the need to consider analysis of disease maps which have a temporal dimension. Several methods have been proposed to handle spatial and spatio-temporal dimensions of disease risk [Bernardinelli et al., 1995, Waller et al., 1997, Knorr-Held et al., 1998, Julian Besag, 1991].

Spatio-temporal models fall in the broad class of Structured Additive Regression (STAR) models. There are a number of approaches that can be used to handle START models but we used a Bayesian approach where all unknown functions and parameters are handled as a unified general framework by assigning an appropriate prior distribution within the same general structure but different forms and degree of smoothness [López-Quılez and Munoz, 2009].

Spatio-temporal models are categorised into three distinct categories according to the temporal evolution of the relative risk in each study region. They can be identified as parametric models, temporal independent models (estimate risk for each period independently of those from the previous periods), and smooth temporal evolution models [López-Quılez and Munoz, 2009]. Under this chapter, we identify and fit a suitable model from each category of spatio-temporal models and map TB prevalence in Kenya for 2002-2009. However, before we embark on model specification and fitting, we look at the general frame work for spatio-temporal modelling.

6.1 Methodology

Consider the case where a given region of interest is divided into N areas (regions, districts, counties or municipalities) indexed by $i = 1, 2, \ldots, n$. Let the temporal dimension be indexed by $t = 1, 2, \ldots, T$, representing each period of time under study. Let n_{it} be the number of persons-times at risk in region i at period t and y_{it} be the corresponding observed cases which are counts [López-Quılez and Munoz, 2009]. Therefore, the best candidate distribution for such a variable is the Poisson distribution.

In this study, we assumed observations y_{it} in region i and period t to be conditionally independent random variables from the exponential family. The observed data Y_{it} depends on N_{it}, the number of people at risk in region i and period t in the study population observed. Let N_{it}^s be the number of people at risk in the standard population, y_{it}^s be the observed TB cases in the standard population and C_{it}^s be the crude rate of TB cases in the standard population. Therefore, the crude rate for region i and period t C_{it}^s is defined by $C_{it}^s = \frac{y_{it}^s}{N_{it}^s}$. It follows that the number of TB cases expected in region i and period t, E_{it} is defined by $E_{it} = C_{it}^s N_{it} = \frac{y_{it}^s}{N_{it}^s} N_{it}$. We used E_{it} as an offset when modelling the TB cases. Therefore, the overall crude rate of TB cases is defined by $\mathbf{C} = \sum_i^n \sum_t^T \frac{y_{it}^s}{N_{it}^s}$ and the overall number of expected TB cases is defined by $\mathbf{E} = \sum_i^N \sum_t^T C_{it}^s N_{it} = \sum_i^N \sum_t^T \frac{y_{it}^s}{N_{it}^s} N_{it}$. We assumed that y_{it} follows the Poisson distribution with expectation $E(y_{it}) = \mu_{it} = E_{it}\vartheta_{it}$, where ϑ_{it} denotes the disease risk in region i, at period t.

We assumed that $y_{it} \sim Poisson\left(E_{it} \exp\left(\eta_{it}\right)\right)$, where $\eta_{it} = \mu + Z_i + A_t + ZA_{it} + u_{it}$ is a linear predictor, μ denotes the grand mean, Z_i the main effect of region i, A_t the temporal trend effect in period t, ZA_{it} is interaction of risk in space and time and u_{it} is the unstructured random effect.

The contribution of a given term may serve to increase or decrease the risk of disease. The intercept or μ gives a background amount of risk shared by all regions and periods.

Most often, an unstructured extra variability term u_{it} is included in the model so as to capture the overall effect of the other unaccounted and unobserved effects. This is often implemented as

a white noise random effect defined as

$$u_e \mid \tau_u^2 \overset{iid}{\sim} N\left(0, \tau_u^2\right), e \in \{i, t, it, \} \tag{6.1.1}$$

Most of the time, smooth and flexible evolutions is preferred for the temporal effect A_t. This term is often modelled as a structured random effect, ensuring that contiguous periods are likely to be similar, but allowing for flexible shape in the evolution curve, particularly when long periods of time are being considered [López-Quılez and Munoz, 2009].

In the following section, we discuss the Bernardinelli et al. [1995] parametric model's approach, Waller et al. [1997] temporally independent spatial model's approach, and the [Knorr-Held, 1999] smooth temporal evolution model's approach. Markov Chain Monte Carlo via Gibbs sampling was used to obtain parameters of interest in each model.

6.2 Bernardinelli et al. (1995) Approach

This is a kind of parametric model in which the log risk in a given region is considered as a linear or quadratic function of time. The function coefficients are region-specific and are spatially structured so that neighbouring regions have similar evolution [López-Quılez and Munoz, 2009]. This model incorporates spatio-temporal interaction where temporal trend in risk may differ for spatial locations and may even have a spatial structure [Lawson et al., 2003]. All temporal trends are assumed to be linear and information is shared in both space and time [Lawson et al., 2003].

To investigate statistical linear rise in the reported TB prevalence in Kenya from 2002 to 2009, we used the Bernardinelli et al. [1995] model.

Assume that $y_{it} \sim Poisson\left(E_{it} \exp\left(\eta_{it}\right)\right)$. According to Bernardinelli et al. [1995], the linear predictor η_{it} is defined by $\eta_{it} = \mu + u_i + v_i + (\varpi + \delta_i) \times \Im_t$, where $u_i + v_i$ follows the BYM specifications [Julian Besag, 1991] (See Section 5.2) and $\varpi \Im_t$ linear trend in time $\Im_t$, δ_i is the interaction random effect between region and time and ϖ is the overall linear time trend. The log relative risk for area i and period t is $\log\left(\vartheta_{it}\right) = \eta_{it}$. Therefore the relative risk of disease is

given by

$$\vartheta_{it} = \exp(\eta_{it}) = \exp(\mu + u_i + v_i + (\varpi + \delta_i) \times \Im_t). \tag{6.2.1}$$

The log of the Poisson mean $\mu_{it} = E_{it} \exp(\mu + u_i + v_i + (\varpi + \delta_i) \times \Im_t)$ is therefore given by

$$\log(\mu_{it}) = \log(E_{it}) + \mu + u_i + v_i + (\varpi + \delta_i) \times \Im_t \tag{6.2.2}$$

6.2.1 Parameter Estimation

Given that $y_{it} \sim Poisson(E_{it} \exp(\eta_{it}))$ with likelihood function denoted by

$$P(\mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta} \mid \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\delta}, \varpi, \tau_u^2, \tau_v^2, \tau_\delta^2), \tag{6.2.3}$$

the prior distribution $P(\boldsymbol{u})$ of $\boldsymbol{u}$ follows a normal distribution (See chapter 4) and prior distribution $P(\boldsymbol{v})$ of $\boldsymbol{v}$ has CAR structure (See chapter 5). Also, δ_i is modelled as a CAR structure with prior distribution denoted by $P(\boldsymbol{\delta})$ and $\varpi \sim N(0, 0.005)$ with prior distribution $P(\varpi)$. The overall mean was defined as $\mu \sim N(0, 0.01)$. Therefore, the posterior distribution is defined as

$$P(\mathbf{u}, \boldsymbol{v}, \boldsymbol{\delta}, \varpi, \tau_u^2, \tau_v^2, \tau_\delta^2 \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta}) \propto P(\mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta} \mid \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\delta}, \varpi, \tau_u^2, \tau_v^2, \tau_\delta^2) P(\boldsymbol{u}) P(\boldsymbol{v}) P(\boldsymbol{\delta}) P(\varpi). \tag{6.2.4}$$

One limitation of the Bernardinelli et al [1995] model is the assumption of a linear time trend in each region. This limitation is resolve by Knorr-Held [2000] model.

Parameters estimation from equation (6.2.4) was carried out using MCMC via Gibbs sampling.

6.3 Waller et al. (1997) approach

This is a type of temporal independent spatial model where spatial effects are simply seen as a set of spatial models, one for each period of time, with almost no relation between them, except possibly for some restriction in their precision parameters. Here temporal evolution is not restricted to any shape and also information is shared in space [López-Quılez and Munoz, 2009]. In this model, the hierarchical specification is applied to each time point separately [Lawson et al.,

2003, Julian Besag, 1991]. This model does not have a single spatial main effect and does allow spatial pattern at each time point to be completely different [Lawson et al., 2003, López-Quılez and Munoz, 2009].

Assume that $y_{it} \sim Poisson\left(E_{it}\exp\left(\eta_{it}\right)\right)$, where $\eta_{it} = u_i^{(t)} + v_i^{(t)}$ and $\mu_{it} = E_{it}\exp\left(\eta_{it}\right)$ is the Poisson mean.

The log relative risk for area i and period t is $\log\left(\vartheta_{it}\right) = \eta_{it}$. Therefore relative risk of disease is given by

$$\vartheta_{it} = \exp\left(\eta_{it}\right) = \exp\left(u_i^{(t)} + v_i^{(t)}\right), \tag{6.3.1}$$

where for each period t, the model term $u_i^{(t)} + v_i^{(t)}$ follows the BYM specification (See Section 5.2) with different precision parameter $\tau_u^{(t)}$ and $\tau_v^{(t)}$ for each period of time. The log of the Poisson mean $\mu_{it} = E_{it}\exp\left(u_i^t + v_i^t\right)$ is therefore given by

$$\log\left(\mu_{it}\right) = \log\left(E_{it}\right) + u_i^t + v_i^t \tag{6.3.2}$$

The $u_i^{(t)}$ and $v_i^{(t)}$ are respectively uncorrelated and correlated heterogeneity terms which may vary with time. This approach results in spatio-temporal model where the spatial dimension is nested within time; thus in effect a spatial model is fitted for each period. The spatial model is not in any way tied or bound to its temporal neighbours, therefore, allowing for free evolution, but not sharing information in time.

This model is used to study the spatial pattern of TB prevalence at each time point in Kenya for $2002 - 2009$.

6.3.1 Parameter Estimation

Given that $y_{it} \sim Poisson\left(E_{it}\exp\left(\eta_{it}\right)\right)$ with likelihood function $P\left(\mathbf{y}, \mathbf{E}, \vartheta \mid \boldsymbol{u}, \boldsymbol{v}, \tau_u^2, \tau_v^2\right)$, the prior distribution $P\left(\boldsymbol{u}\right)$ of $\boldsymbol{u}$ follows a normal distribution (See chapter 4) and prior distribution $P\left(\boldsymbol{v}\right)$ of $\boldsymbol{v}$ has CAR structure (See chapter 5). Therefore, the posterior distribution is defined as

$$P\left(\boldsymbol{u}, \boldsymbol{v}, \tau_u^2, \tau_v^2 \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta}\right) \propto P\left(\mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta} \mid \boldsymbol{u}, \boldsymbol{v}, \tau_u^2, \tau_v^2\right) P\left(\boldsymbol{u}\right) P\left(\boldsymbol{v}\right). \tag{6.3.3}$$

Estimation of parameters from equation (6.3.3) was achieved through Bayesian MCMC via Gibbs sampling.

6.4 Knorr-Held and Rasser, 2000

This is a type of smooth temporal evolution model where the evolution of the estimated risk in each region is a smooth function of time. Knorr-Held [1999] proposed this model to overcome the limitation suffered by the Bernardinelli et al. [1995] model.

Assume that $y_{it} \sim Poisson\left(E_{it} \exp\left(\eta_{it}\right)\right)$. Knorr-Held [1999] defined the linear predictor η_{ij} of a nonparametric, additive model as

$$\eta_{it} = \mu + u_i + v_i + \Im_t + \psi_{it}, \tag{6.4.1}$$

where the model term $u_i + v_i$ follows the BYM specification. The parameter $\Im_t$ represents an unstructured or structured temporal effect and the parameter ψ_{it} is the space-time interaction. The log relative risk for area i and period t is $\log\left(\vartheta_{it}\right) = \eta_{it}$. Therefore the relative risk of disease is given by

$$\vartheta_{it} = \exp\left(\mu + u_i + v_i + \Im_t + \psi_{it}\right). \tag{6.4.2}$$

The log of the Poisson mean $\mu_{it} = E_{it} \exp\left(\mu + u_i + v_i + \Im_t + \psi_{it}\right)$ is therefore given by

$$\log\left(\mu_{it}\right) = \log\left(E_{it}\right) + \mu + u_i + v_i + \Im_t + \psi_{it}. \tag{6.4.3}$$

It should be noted that $\boldsymbol{u}, \boldsymbol{v}$ and $\Im$ are the main effects whiles $\boldsymbol{\psi}$ is the space-time interaction term.

This model is used to study smooth temporal evolution of the estimated relative risk of TB prevalence in Kenya in each region at given point in time.

6.4.1 Parameter Estimation

Given that $y_{it} \sim Poisson\left(E_{it} \exp\left(\eta_{it}\right)\right)$ with likelihood function

$$P\left(\mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta} \mid \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\psi}, \boldsymbol{\Im}, \tau_u^2, \tau_v^2, \tau_\psi^2, \tau_\Im^2\right), \tag{6.4.4}$$

the prior distribution $P\left(\boldsymbol{u}\right)$ of $\boldsymbol{u}$ has a normal distribution (See chapter 4) and prior distribution $P\left(\boldsymbol{v}\right)$ of $\boldsymbol{v}$ has CAR structure (See chapter 5). According to Knorr-Held [1999], if the main temporal random effect $\Im_t$ assumes unstructured random effect, then its prior distribution would be $\Im_t \mid \tau_\Im^2 \sim N\left(0, \tau_\Im^2\right)$ and if it assumes structured random effect, then its prior density follows a first order random walk defined by

$$P\left(\boldsymbol{\Im} \mid \tau_\Im^2\right) \propto \exp\left[-\frac{\tau_\Im^{-2}}{2} \sum_{t=2}^{T} \left(\Im_t - \Im_{t-1}\right)^2\right]. \tag{6.4.5}$$

The second-order random walk is also possible.

According to Knorr-Held [1999], prior specification for the interaction term $\boldsymbol{\psi}$ depends on the spatial and temporal main effect which are assumed to interact. Different types interactions ψ_{it} were classified by Knorr-Held [1999] with prior distribution denoted by $P\left(\boldsymbol{\psi}\right)$ and precision variance denoted by τ_ψ^2. Therefore, the posterior distribution for the relative risk $\boldsymbol{\vartheta}$ is defined by

$$P\left(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\psi}, \boldsymbol{\Im}, \tau_u^2, \tau_v^2, \tau_\psi^2, \tau_\Im^2 \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta}\right) \propto P\left(\mathbf{y}, \mathbf{E}, \boldsymbol{\vartheta} \mid \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\psi}, \boldsymbol{\Im} \tau_u^2, \tau_v^2, \tau_\psi^2, \tau_\Im^2\right) \times \tag{6.4.6}$$

$$P\left(\boldsymbol{u}\right) P\left(\boldsymbol{v}\right) P\left(\boldsymbol{\psi}\right) P\left(\boldsymbol{\Im}\right). \tag{6.4.7}$$

The interaction type depends on which of the two possible type of temporal effects (unstructured or structured) interacts with the two main effects (u_i and v_i). Each of the four type of interactions has different prior interrelationships involving the interaction term ψ_{it} [Knorr-Held, 1999].

1. **Interaction type I**: If the unstructured main effects ($\Im_t$ and u_i) are expected to interact, then the distribution of the interaction parameter ψ_{it} is defined as

$$P\left(\psi \mid \tau_\psi\right) \propto \exp\left[-\frac{\tau_\psi}{2} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\psi_{it}\right)^2\right]. \tag{6.4.8}$$

   This may be considered as an independent unobserved covariate for each combination of region and period (i, t), thus without any structure [Knorr-Held, 1999, López-Quılez and

Munoz, 2009]. On the other hand, if spatial and temporal main effects are present in the model, then the interaction effect only denote independence in the deviations from them. The main effects can cause contribution to risk in neighbouring regions or in consecutive period of time to be highly correlated. This is a global space-time heterogeneity effect and it is often modelled as white noise defined as $\psi_{it} \sim N\left(0, \tau_{\psi}^{2}\right)$. This interaction type has independent prior with no structure in space-time interaction [Knorr-Held, 1999, López-Quılez and Munoz, 2009].

2. **Interaction type II**: Knorr-Held et al. [1998] noted that this interaction effect is distributed as a random walk independently of other counties if we modelled $\Im_t$ as a random walk. The prior distribution for this interaction is defined by [Knorr-Held et al., 1998, Knorr-Held, 1999, Lawson, 2008]

$$[\psi \mid \tau_{\psi}] \propto \exp\left[-\frac{\tau_{\psi}}{2}\sum_{i=1}^{n}\sum_{t=2}^{T}\left(\psi_{it}-\psi_{i,t-1}\right)^{2}\right] \tag{6.4.9}$$

This type of interactions has no structure in space [Knorr-Held et al., 1998]. This implies that each region has a specific evolution structure that is independent of that in the neighbouring region [López-Quılez and Munoz, 2009, Knorr-Held, 1999].

3. **Interaction type III**: If we assumed that the unstructured temporal main effect ($\Im_t$) and the spatially correlated or structured main effect (υ_i) interact, then the interaction effect parameter $\psi_t = (\psi_{1t}, \ldots, \psi_{nT})', t = 1, \ldots, T$ follows an independent Intrinsic autoregressive distribution defined as [Knorr-Held, 1999, Knorr-Held et al., 1998, Lawson, 2008]

$$[\psi \mid \tau_{\psi}] \propto \exp\left[-\frac{\tau_{\psi}}{2}\sum_{t=1}^{T}\sum_{i\sim\ell}\left(\psi_{it}-\delta_{lt}\right)^{2}\right] \tag{6.4.10}$$

This interaction is assumed to have a spatial structure for each period, independent of adjacent periods (its neighbours in time). This interaction type is analogous to the clustering effect, which is often modelled as a CAR distribution (section 5.1) for each period [López-Quılez and Munoz, 2009, Knorr-Held et al., 1998]. Here we implicitly assumed that each specific region may have a slight deviation from the global trend, but that this deviation is likely to be identical to that in the neighbouring regions, while at the same time,

independent of that in that in the previous period of time [López-Quılez and Munoz, 2009, Knorr-Held, 1999].

4. **Interaction type IV**: Type IV is completely dependence over space and time theoretically [López-Quılez and Munoz, 2009, Knorr-Held et al., 1998]. Hence the effect can no longer be factorized into independent blocks if $\Im_t$ is modelled as a random walk allowing interaction with the structured main effect (υ_i). Knorr-Held [1999] defined type IV interaction as

$$[\psi \mid \tau_\psi] \propto \exp\left[-\frac{\tau_\psi}{2}\sum_{t=2}^{T}\sum_{i\sim\ell}(\psi_{it}-\psi_{lt}-\psi_{i,t-1}+\psi_{l,t-1})^2\right] \tag{6.4.11}$$

Knorr-Held [1999] stated that, type IV is the most interesting type of interaction that occurs when deviation from the global trends are highly correlated with their neighbours, both in space and time. Here, hidden factors whose effects exceed the limits of one or more regions and also persistent for one or more period of time can be modelled. This is also an efficient way of obtaining information from data, particularly in the case of rare diseases or less populated regions, since the risk estimation for the region-period is not performed on the basis of only locally observed data but also on that in the neighbouring regions and periods [López-Quılez and Munoz, 2009].

The hyperprior distribution for $\tau^2_\Im$ and τ^2_ψ are modelled as gamma distribution. Knorr-Held (2000) fitted four different types of interaction effects to the 21 years Ohio respiratory cancer dataset and found that interaction type II was appropriate; offering lowest deviance [Lawson et al., 2003].

Depending on the data you are dealing with, any of the four interaction types can yield best fit for the data [Knorr-Held, 1999]. Spatio-temporal Parametrisation of log relative risk can take a variety of forms and it is not clear yet which form is most appropriate [Lawson et al., 2003].

Estimation of all parameters was achieved with Bayesian MCMC via Gibbs sampling.

6.5 Application of Spatio-Temporal Models

Best fitting spatio-temporal model to Kenya TB data was selected from the above candidate models based on their respective model's DICs and pDs presented in Table 6.5

Table 6.1 the presents results of fitting the spatio-temporal models to Kenya TB data. Note that these values are subject to Monte Carlo error, which is difficult to quantify. We have therefore chosen a very long run of which convergence was reached at 1,200,200 after a burn-in period of 100,000 and thinning of every 30^{th} element of the chain for each model.

Table 6.1: Spatio-Temporal models Deviance Summaries

Model indicators	Bernardinelli et al., 1995	Waller et al.,1997	Knorr-Held et al., 20000
$\bar{D}$	135492	4039.320	3818.720
pD	487.536	594.296	375.306
DIC	135979	4633.62	4194.03

From Table 6.2, $\bar{D}$ is the mean of the posterior deviance, pD is the effective number of parameters and DIC $= \bar{D}+$ pD proposed by Spiegelhalter et al. [2002].

Among the spatio-temporal models presented in Table 6.1, the model with the lowest DIC (4194.30) and lowest pD (375.306) is the Knorr-Held et al [2000] model, equation (6.4.3). We therefore recommend equation (6.4.3) model as the best fitting space-time model to Kenya TB data for $2002 - 2009$.

We now evaluate all the four possible types of interactions discussed in Section 6.1. Table 6.2 summaries our effort to identify a best fitting model.

Table 6.2 presents deviance summary of the interaction types after MC convergence at 1,200,000 and a burn-in period of 100,000 for each model. The model fit with interaction type III and IV fit the data well but type IV seems better than type III since it yields the lowest pD (362.494) and DIC (419.410). We now provide a more detailed look at the results of the type IV model as it gives best fit with less posterior deviance.

Table 6.2: Interaction Types Deviance Summaries

Model indicators	Type I	Type II	Type III	Type IV
$\bar{D}$	3820.060	3818.600	3827.920	3826.910
pD	376.637	374.377	363.510	362.494
DIC	4196.700	4192.970	4191.430	4189.410

Table 6.3: Interaction Type IV Posterior Statistics

Model indicators	estimates	95% Credible Interval
μ	-0.22	(-0.54,0.70)
τ_v^2	10.8	(1.39,50.10)
τ_u^2	9.17	(3.57,28.60)
τ_ψ^2	11.3	(9.55,13.20)

In Table 6.3, the overall mean relative risk μ is insignificant and the precision variance parameters τ_v^2 and τ_u^2 indicates significance of clustering and heterogeneity of relative over the studied period respectively. The precision variance parameter indicates significance of TB relative risk interaction in space-time. We now present maps of the interaction type IV in section 6.5.1

6.5.1 Type IV interaction Model

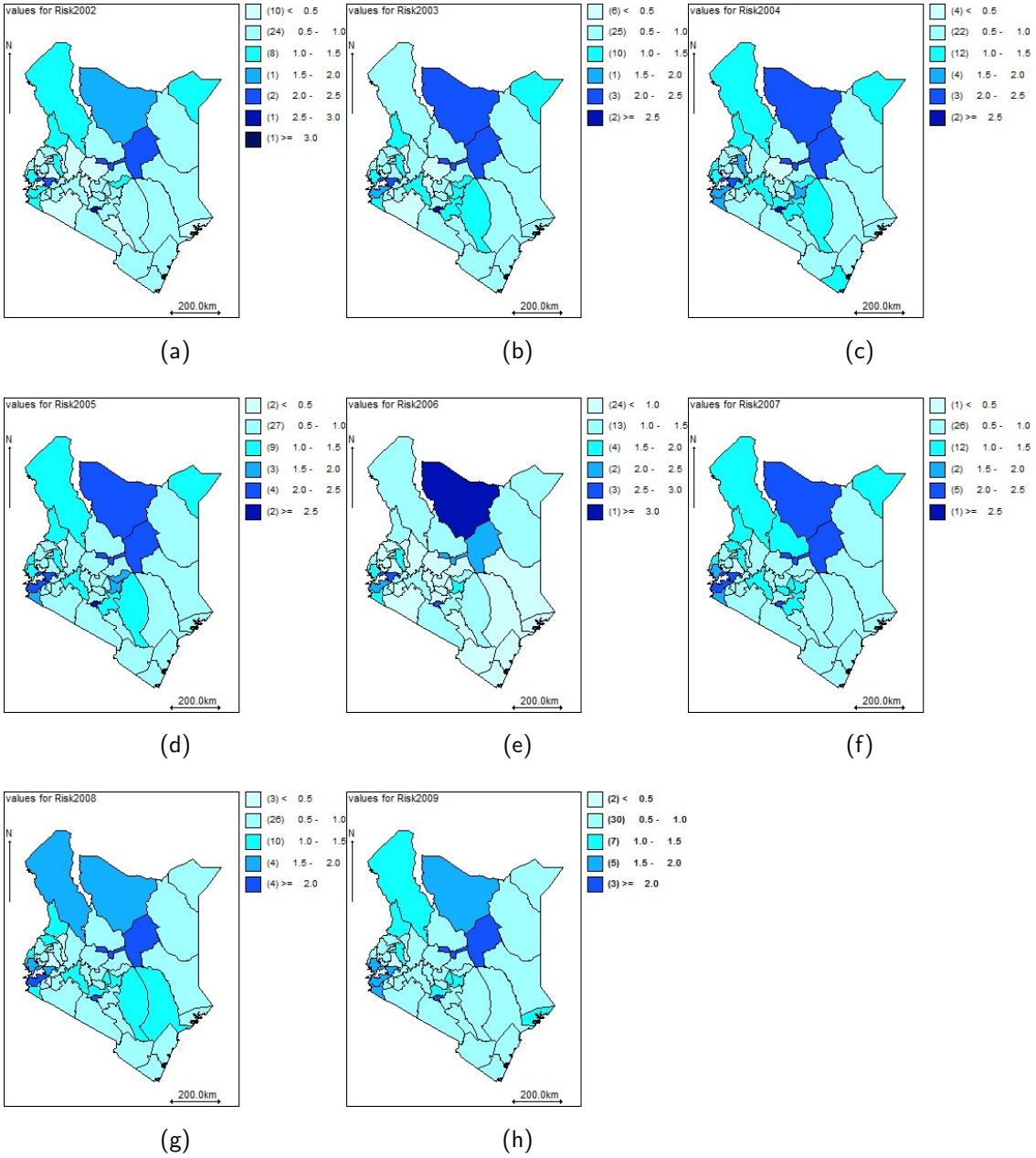


Figure 6.1: Type IV interaction posterior mean of the relative risk maps for 2002-2009.

Figure 6.1 displays the spatial distribution of the posterior relative risk for 2002-2009. Generally, the spatial pattern does not change much over the study period. However, some counties have interesting time trends, for instance, the two adjacent counties (Nairobi and Kiambu) in the

central part of Kenya where opposite trend in disease risk can be detected (See Figure 6.2). This may be due to the fact that type IV interaction borrows strength from neighbouring counties, hence, the decreasing trend in Nairobi county causes the estimated increase in Kiambu which is less populated than Nairobi, to be less pronounced. Again, high risk of TB prevalence is observed in the North, West, North-West and the Central counties and low risk in the South-East counties for 2002-2009.

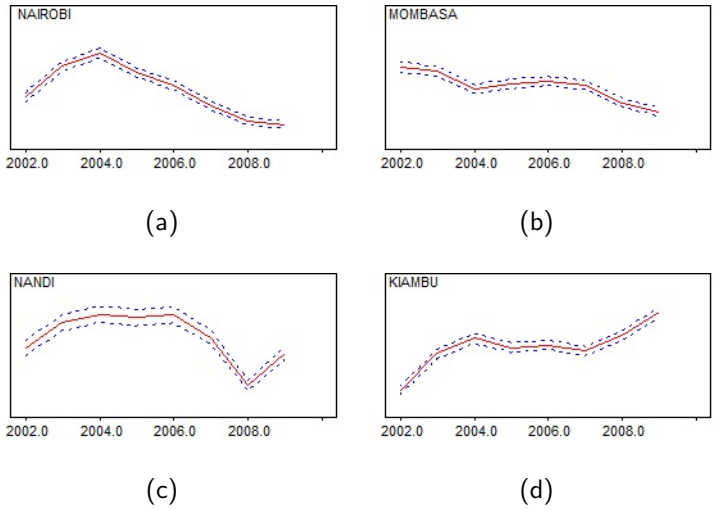


Figure 6.2: Type IV interaction temporal trend Posterior mean of the relative risk 2002-2009

Figure 6.2 displays decreasing temporal trend of posterior relative parameters for some highly urbanized counties such as Mombasa and Nairobi. In contrast, pronounced increasing trends were observed for most rural counties such as Nandi and Kiambu. More information on temporal trend behaviour of posterior relative risk for the rest of the counties can be found in Appendix 7.

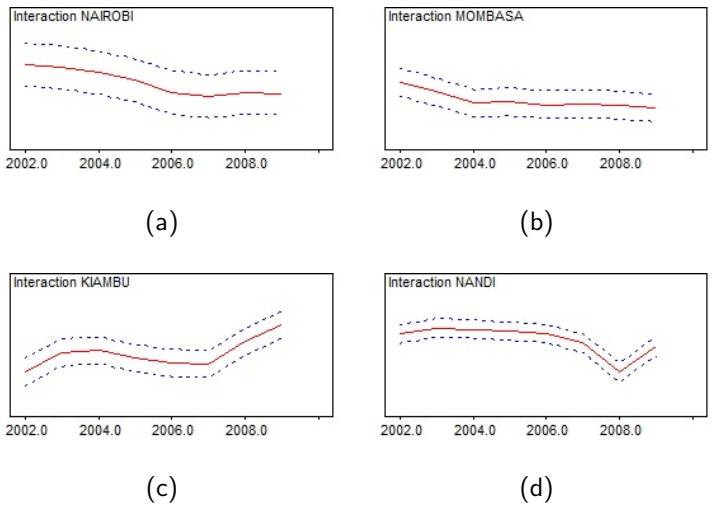


Figure 6.3: Type IV interaction trend of TB prevalence rate in Kenya from 2002-2009

Figure 6.3 displays decreasing temporal trend of interaction parameters for some highly urbanized counties such as Mombasa and Nairobi. For both counties, 95% simultaneous credible regions for ψ_{ij} shows significance of interaction. In contrast, pronounced increasing trends were observed for most rural counties such as Nandi and Kiambu. More interaction trend behaviour of risk in the rest of the counties can be found in Appendix 7.

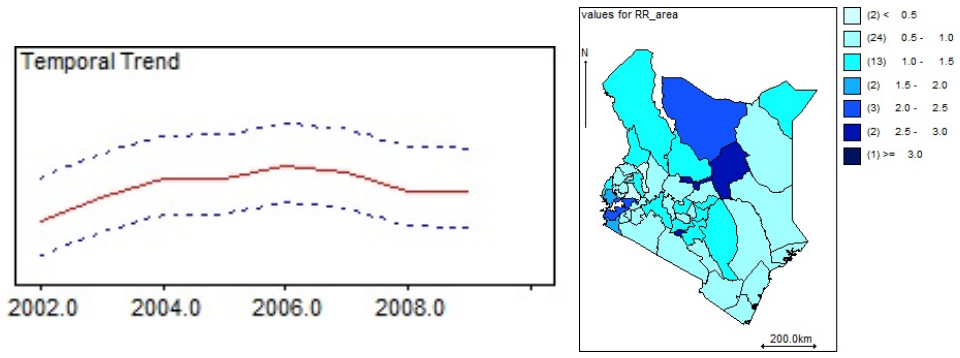


Figure 6.4: Type IV temporal trend and area posterior mean relative risk.

Figure 6.4a displays a slight increasing temporal trend from 200-2004, slight decrease from 2004-2005, increase in 2006, a slight decrease in 2007-2008 and slight increase in 2009. The temporal trend effect does not change much for 2002-2009. Figure 6.4b displays area relative risk of the

type IV interaction. Again high risk of TB is observed in the North, West, North-West and Central counties of Kenya and low risk in the South-East counties.

Interaction type IV was identified by Schrödle and Held [2011] as the best fitting spatio-temporal model for modelling and mapping salmonellosis counts in cattle in Switzerland, 1991-2008.

6.6 Summary of Spatio-Temporal Models

All spatio-temporal models were applied to Kenya TB data during the years 2002-2009. Using the DIC, the best model was chosen and conclusion concerning the pattern of TB prevalence in Kenya drawn. Knorr-Held [1999] parametric model was selected as the best fitting model to Kenya TB data since it yields the lowest DIC. Knorr-Held [1999] parametric model with interaction type IV provides best fit for Kenya TB data.

Interaction of TB relative in space and time is decreasing in most urban counties and increasing in most rural counties. This is due to the fact that type IV model borrow strength from neighbouring counties such that these have similar risk as observed between Nairobi and Kiambu and between Mombasa and Kwale. The temporal trend effect does not change much for 2002-2009

Chapter 7

Discussion and Conclusion

This thesis provides a framework for non-spatial, spatial and spatio-temporal models used in disease modelling and mapping. Table 7.1 presents comparison of the Non-spatial and Spatial Models used in this study.

Table 7.1: 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

Table 7.1 presents overall posterior statistics of the non-spatial and spatial models.

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 confirmed that with all determinants of TB kept constant, overall relative risk of TB would be decreasing. Also, the UH, CAR, and BYM confirmed HIV as a major TB determinant and that TB prevalence in Kenya increases with increasing HIV.

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 neighbouring counties. Among the LN, the UH, the CAR and the BYM models, the UH model yields 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 implies significant difference while a DIC less than 5 implies a negligible difference [Best, 2011], one can use any of the nonspatial and spatial models for fitting Kenya TB data for 2002-2009 depending on the issue at hand.

We have considered several formulations for analysis of spatial and spatio-temporal disease data. The spatial models provide an over view of risk behaviour in space whiles the spatio-temporal models provide an over view of risk behaviour in both space and time. We have also considered several formulations of spatio-temporal disease models in the presence of space-time interaction. In spatio-temporal models, the main effects are combined with interaction parameters. Models can be simplified if interaction turn out to be negligible, else, we examine the posterior distribution of the parameters so as to identify pattern of disease variation which cannot be attributed to the main effects. The Bayesian credible interval for the interaction parameters have been useful to identify those counties which do not follow the overall temporal trend.

These models were fitted with Kenya's TB prevalence data for 2002-2009. Markov Chain Monte Carlo via Gibbs sampling was used for simulation of parameters from posterior distributions. Rubin and Gelman convergence diagnostics test was used to confirm convergence of the Markov Chain. Thinning the Markov Chain and the over-relax algorithm though slow the speed of the MCMC but significantly reduces autocorrelation and number of iterations. Long-run MCMC iterations and high thinning sample size k is require for spatio-temporal models used in fitting Kenya's TB data.

The DIC of each model were compared to select the best model from the set of candidate models used in fitting Kenya's TB prevalence data.

Among the determinants considered, HIV is identified as a major determinant of TB. This finding is consistent with the expectation of increases in HIV prevalence with increases in TB prevalence. Variation in TB risk is observed among Kenya counties and clustering among counties with high TB relative risk. Risk of tuberculosis doe not changes significantly over the study period.

Interaction of TB relative risk in space and time is decreasing in most urban counties and increasing in most rural counties. This is due to the fact that type IV model borrows strength from neighbouring counties such that neighbouring counties have similar risk as observed between.

Unknown or unobserved factors are presumably described by the space and time interaction term ψ_{ij}. The interaction component of the spatio-temporal models is an important aspect in modelling disease risk in space and time.

Generally clustering of risk and elevated risk is observed in the North, West, North-West and the central counties of Kenya and low clustering and elevated risk in the South-West counties.

We have discovered an interesting association between temporal trends of interaction parameters and urbanization in Kenya, which might set a framework for further epidemiological research.

Modelling of risk in space and time is quite a challenging task. Although these approaches are less than ideal, we hope that our formulations provide a useful stepping stone into the development of spatial and spatio-temporal methodology for modelling and mapping Kenya's TB prevalence data.

We are satisfied that the models selected in this thesis are from an appropriate class that led to the analysis of the Kenya's TB data for 2002-2009.

Further research is required for a standard or acceptable distribution type for space-time interaction ψ_{ij} to be identified since comparing posterior Deviance from interaction type that assumed t_j to be modelled as structured or structure could cause one or more deficiencies to a given interaction type.

The limitation of the study is the specification of the adjacency matrix $\mathbf{W}$ with 0 and 1 in

the CAR model is not internally consistent in a case in which the number of neighbours 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 require less computation time but presents theoretical and conceptual issues that undermines its validity [Wang and Kockelman, 2013]. For instance, the precision parameter $\boldsymbol{\tau^2}$ is unknown (which is always the case), the functional from of the joint distribution of the spatial random effects ($\boldsymbol{v}$), are not identifiable under the "intrinsic" CAR specification. Thus one cannot be confident that his/her estimates, nor convergence of the parameters draws, due to potentially improper distributional assumptions. Conceptually, not including ρ in the model blurs one's estimates and can lead to counter-intuitive interpretation [Wang and Kockelman, 2013].

Appendix I: Convergence Diagnostics of the BYM model

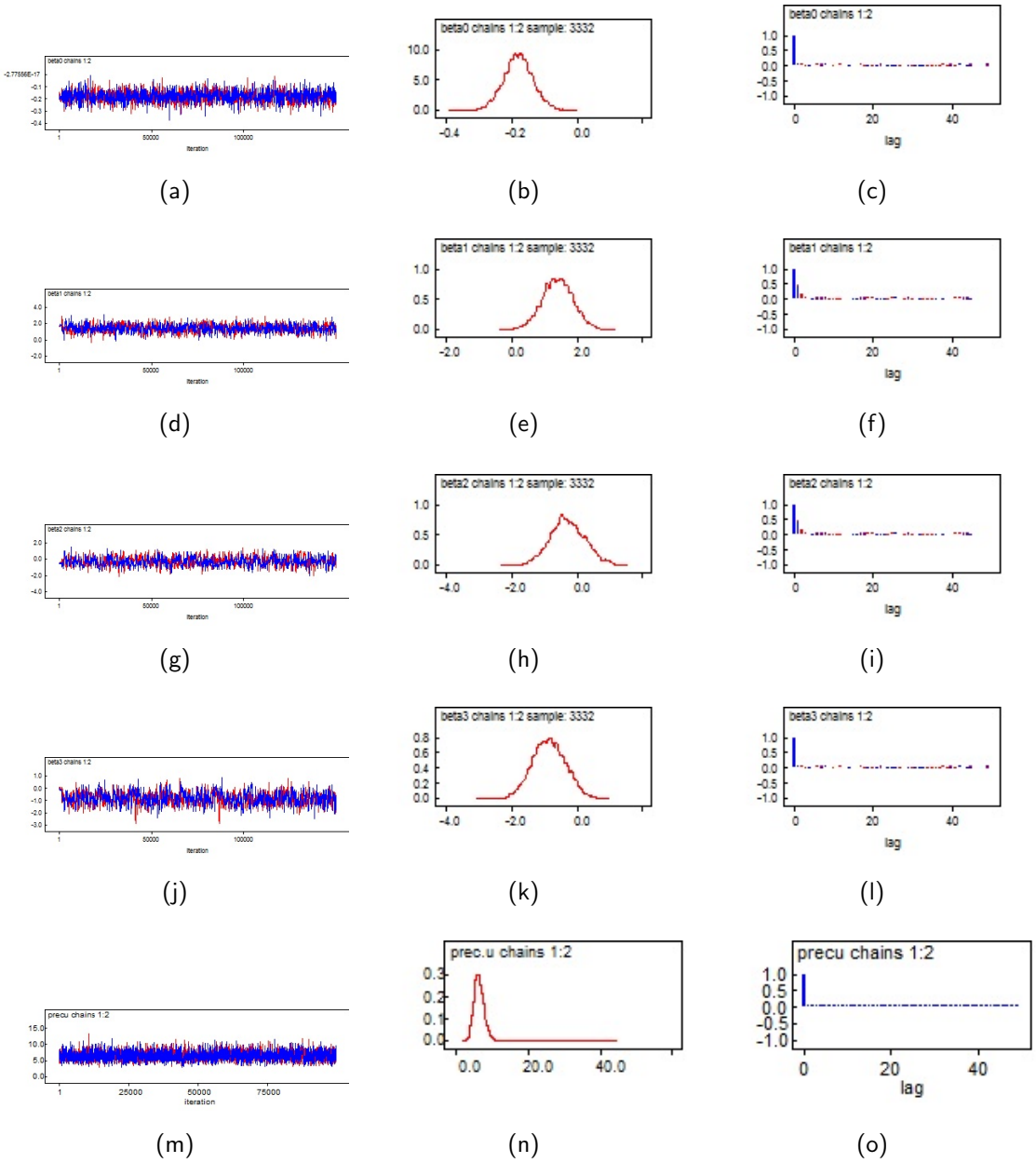


Figure 7.1: BYM Model: Rubin and Gelman Convergence Diagnostics

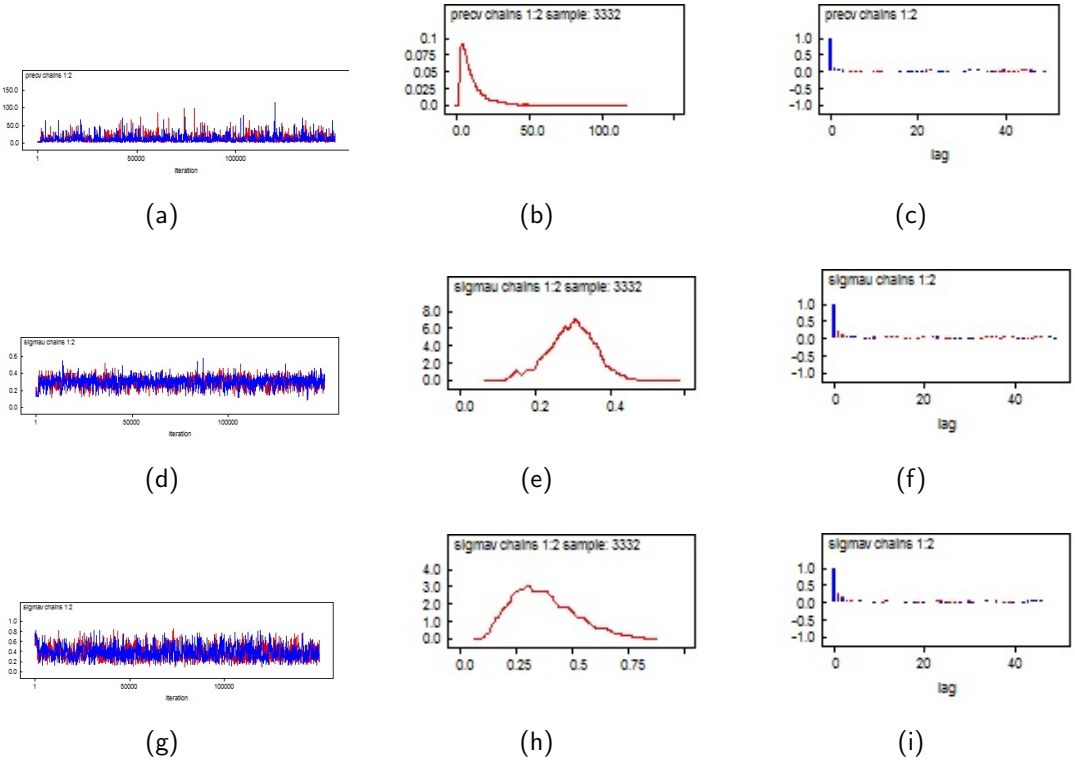


Figure 7.2: BYM model: Rubin and Gelman Convergence Diagnostics

Appendix II: The Interaction Type IV Model's Results

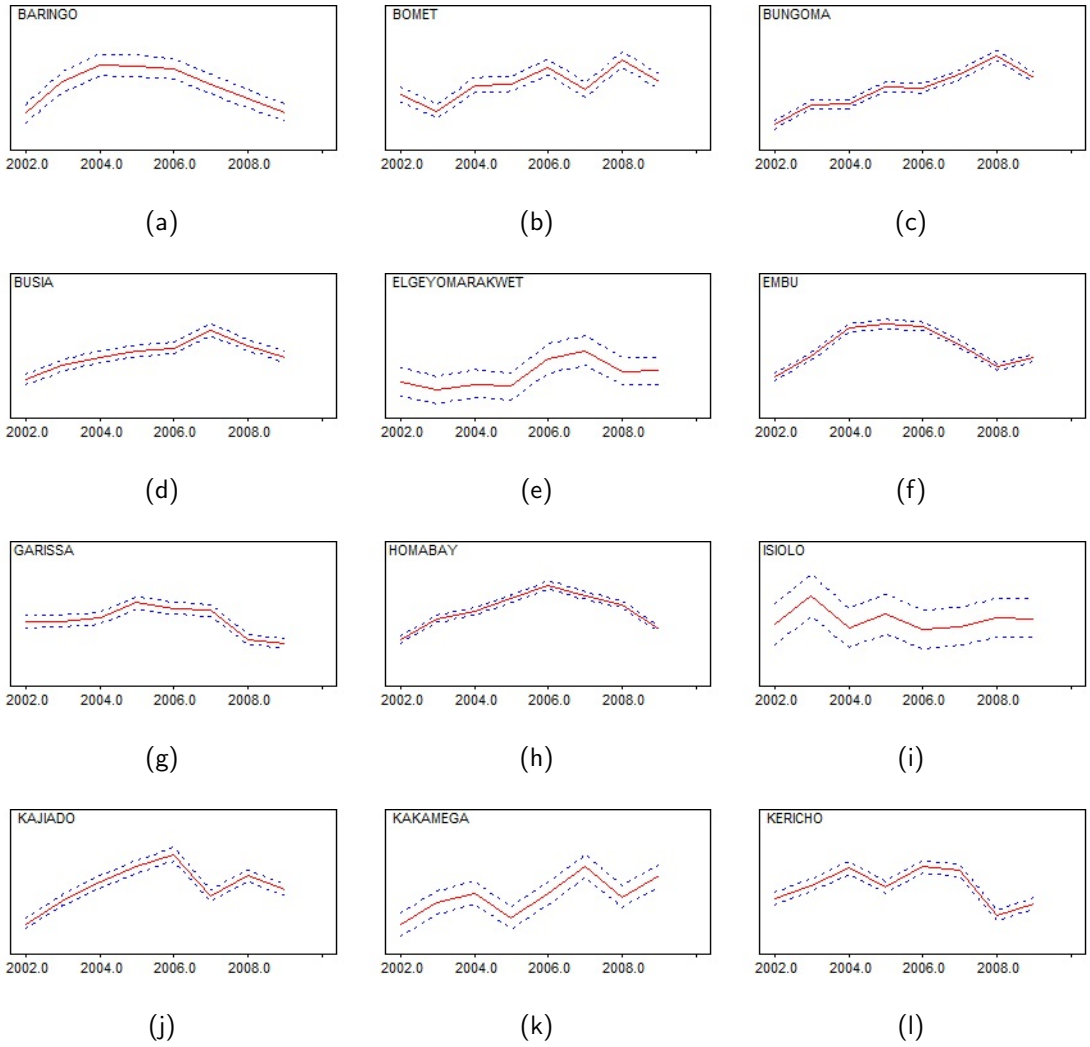


Figure 7.3: Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009

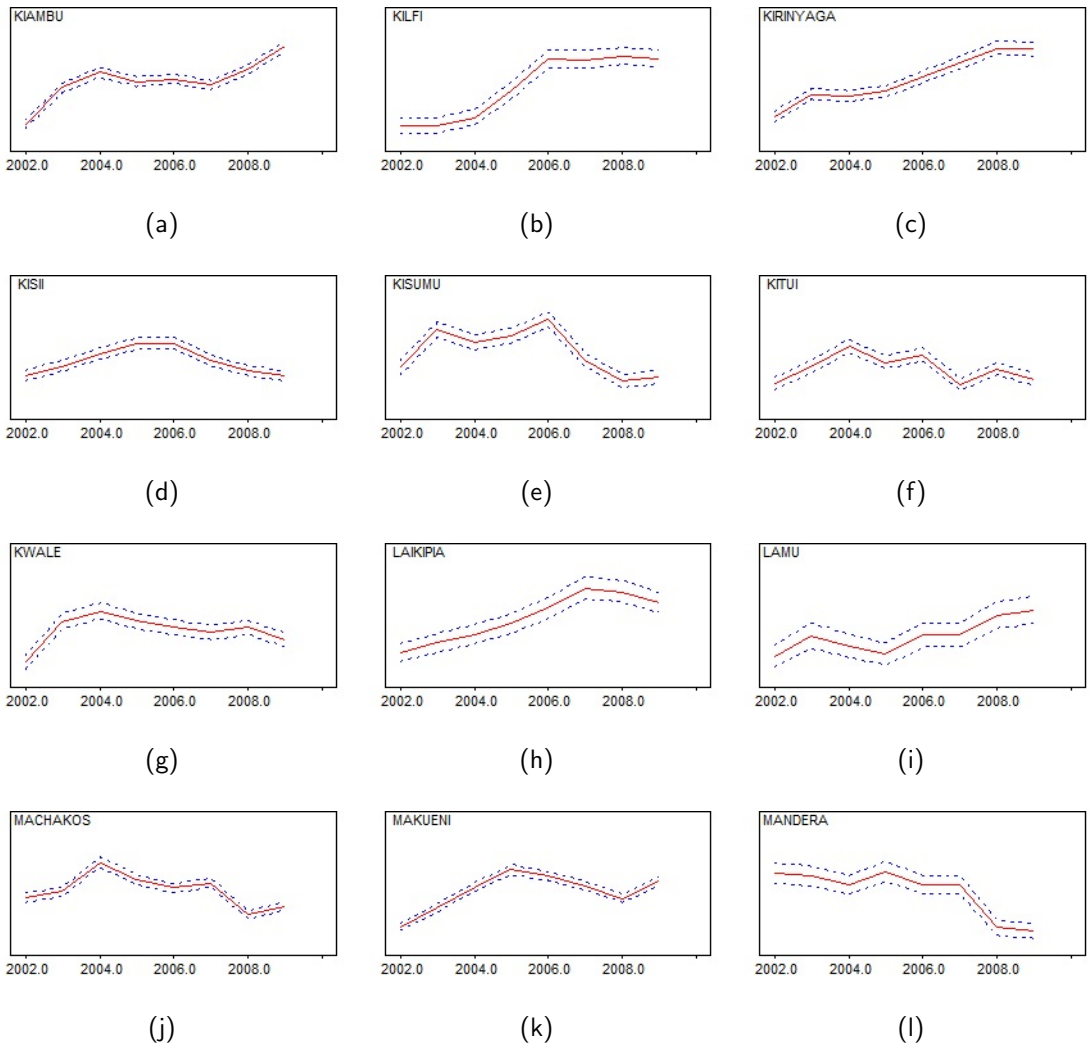


Figure 7.4: Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009 continue

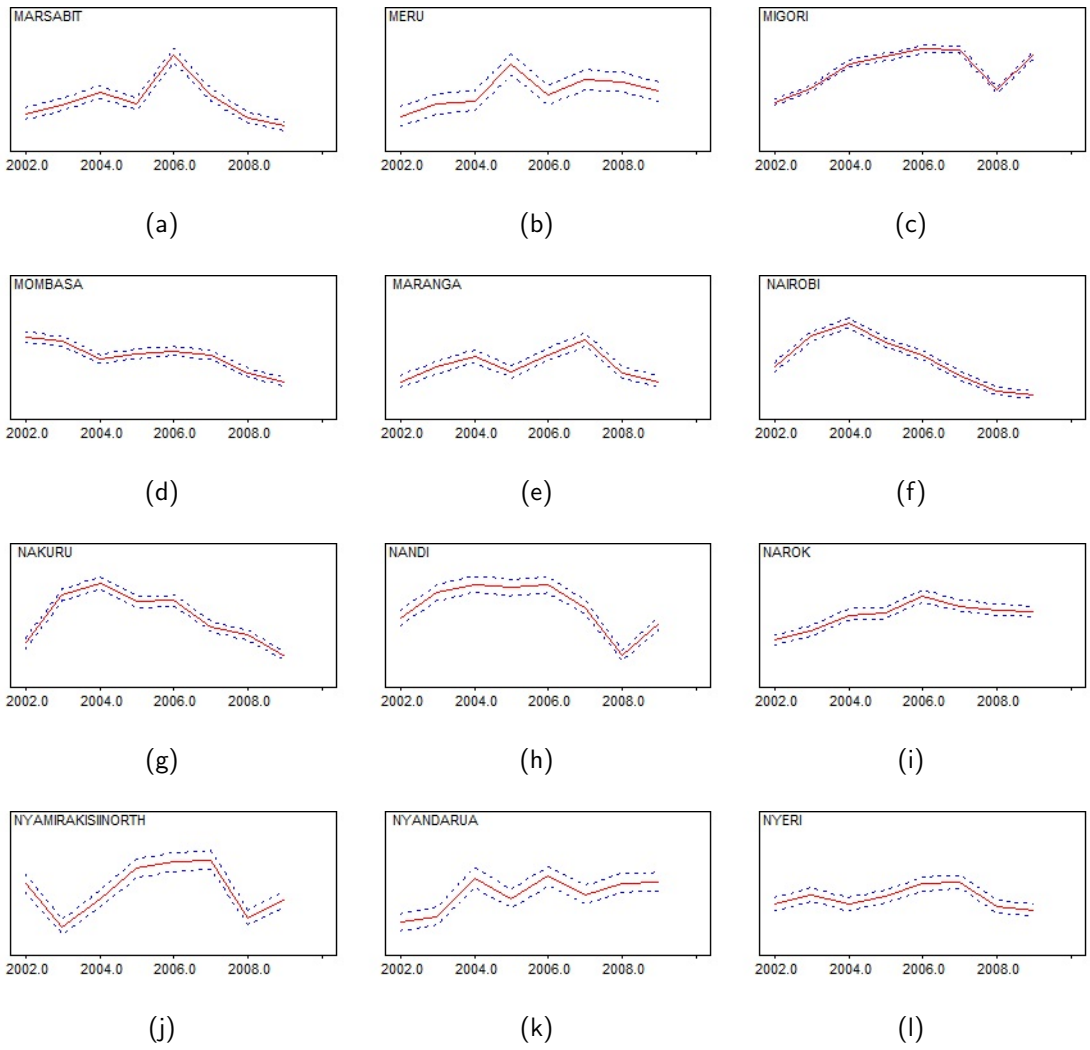


Figure 7.5: Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009 continue

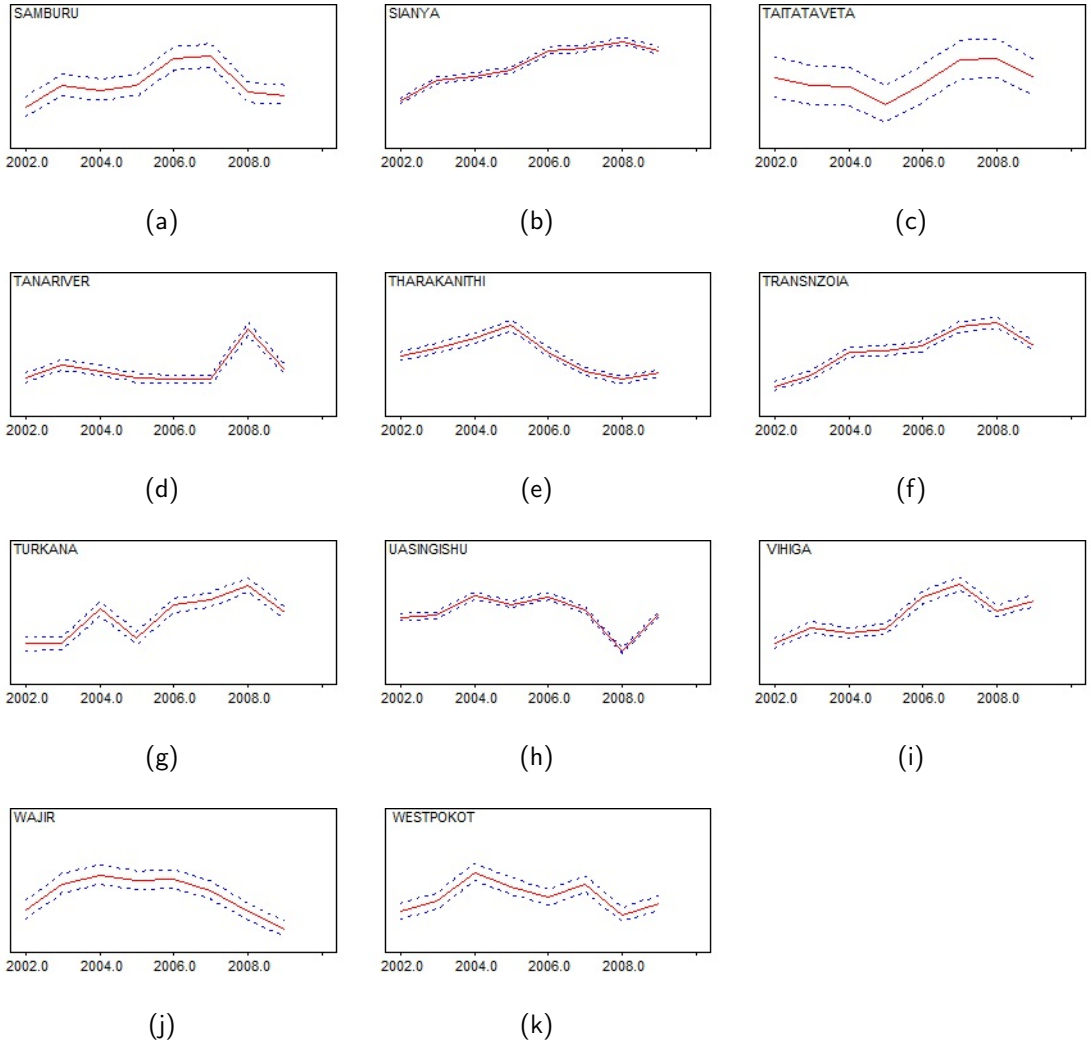


Figure 7.6: Type IV posterior mean relative risk temporal trend of TB prevalence rate in Kenya from 2002-2009 continue

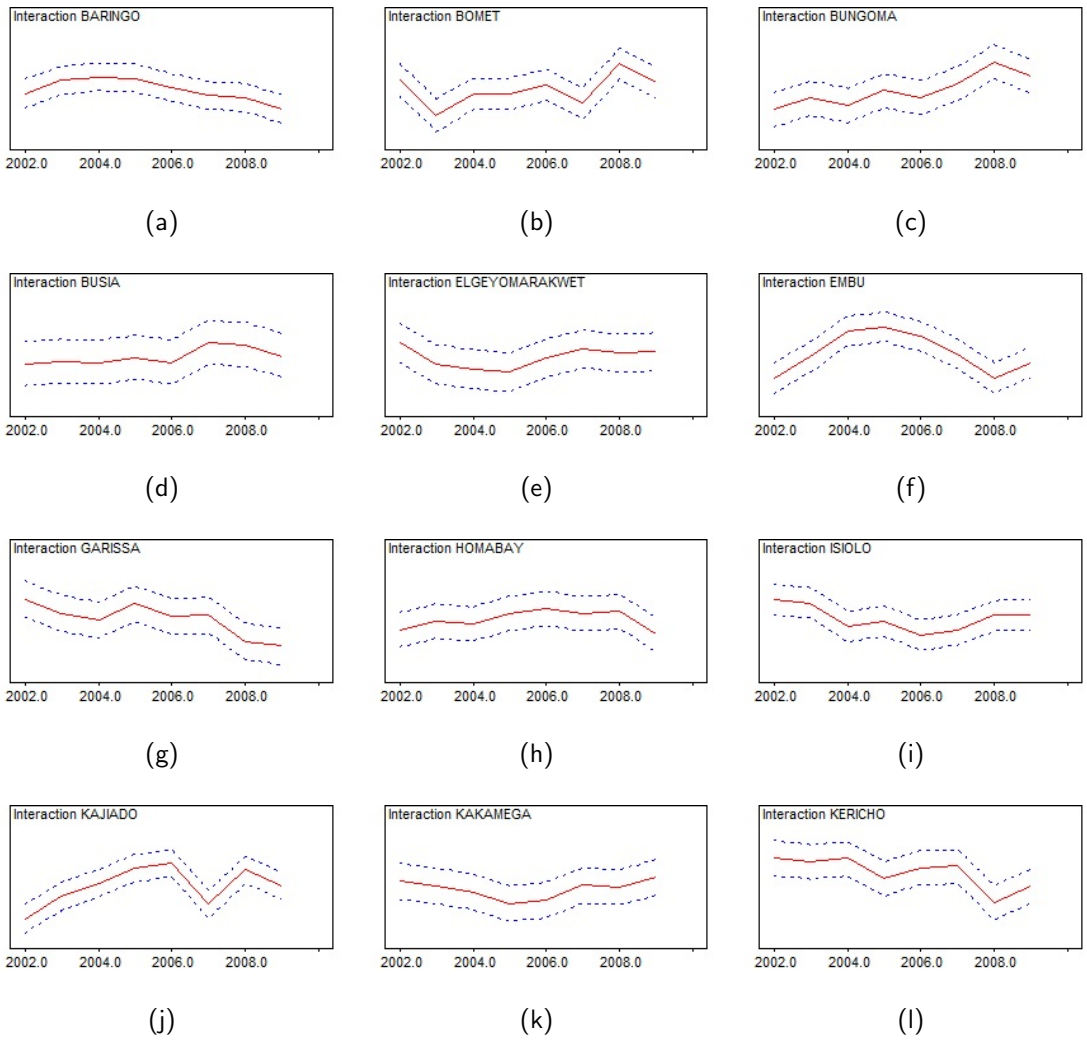


Figure 7.7: Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009

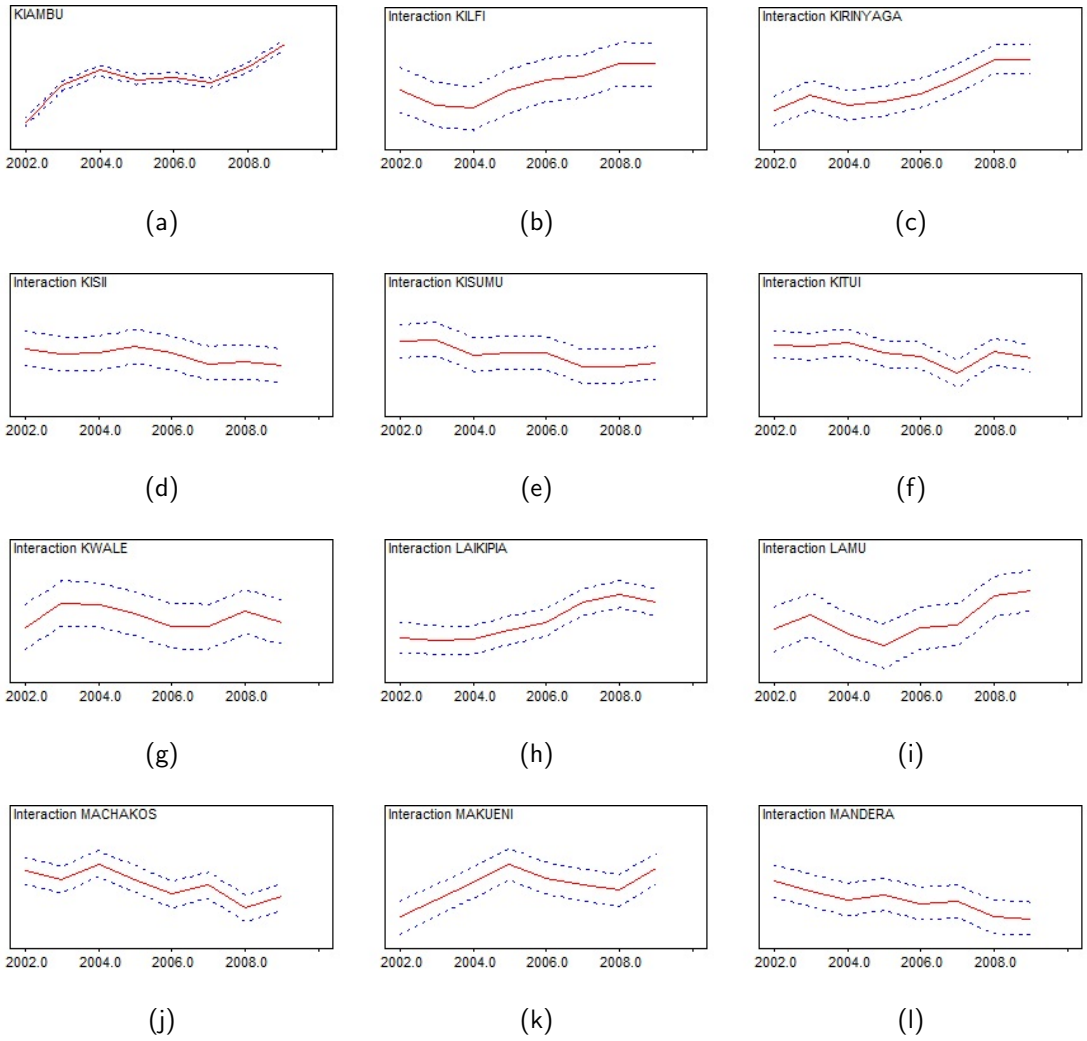


Figure 7.8: Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009 continue

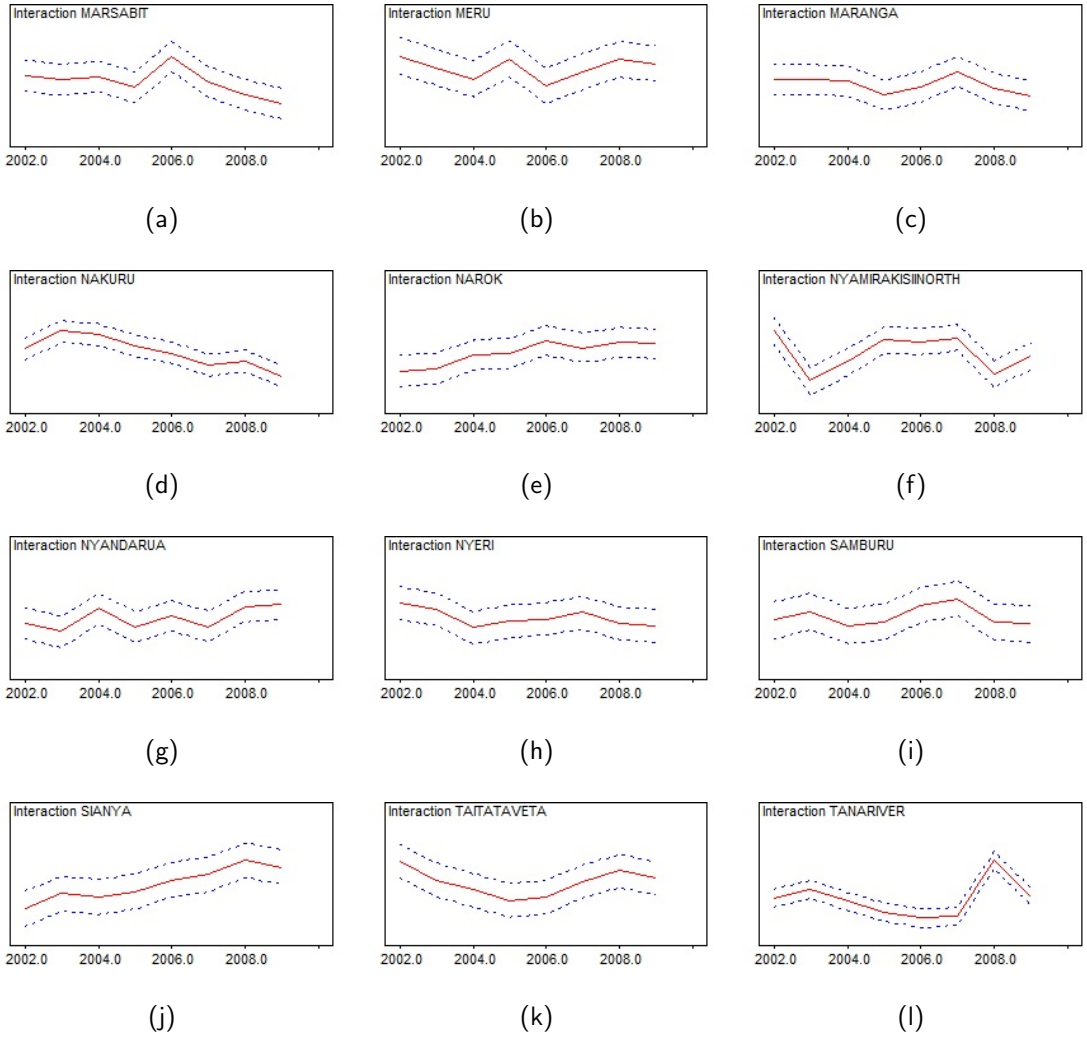


Figure 7.9: Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009 continue

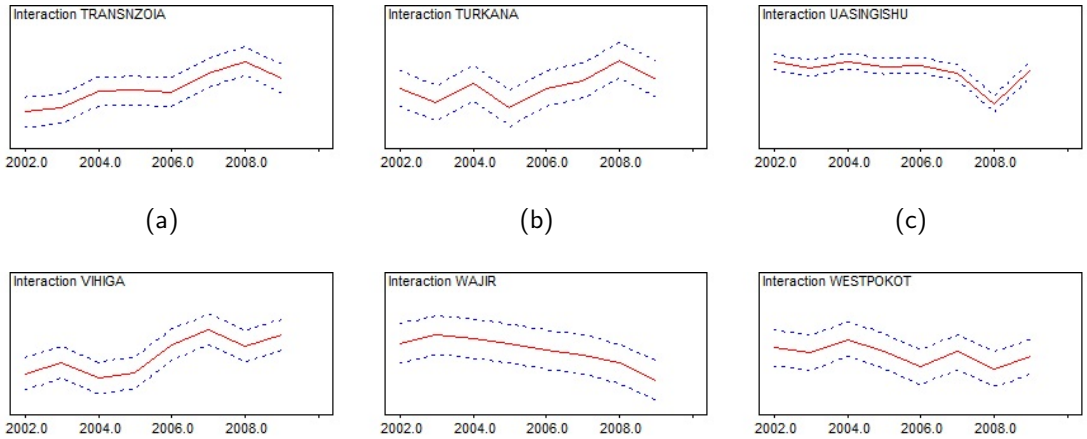


Figure 7.10: Type IV interaction temporal trend of TB prevalence rate in Kenya from 2002-2009 continue

References

Banerjee A, AD Harries, T Nyirenda, and FM Salaniponi. Local perceptions of tuberculosis in a rural district in malawi. *The International Journal of Tuberculosis and Lung Disease*, 4(11): 1047–1051, 2000.

Achcar, EZ Martínez, A Ruffino-Netto, JA, CD Paulino, and P Soares. A statistical model investigating the prevalence of tuberculosis in new york city using counting processes with two change-points. *Epidemiology and infection*, 136(12):1599, 2008.

Hirotugu Akaike. Likelihood of a model and information criteria. *Journal of econometrics*, 16(1): 3–14, 1981.

Christian Auer, Jesus Sarol Jr, Marcel Tanner, and Mitchell Weiss. Health seeking and perceived causes of tuberculosis among patients in manila, philippines. *Tropical medicine & international health*, 5(9):648–656, 2000.

Latouche Aurélien, Guihenneuc-Jouyaux Chantal, and Hémon Denis. Robustness of the bym model in absence of spatial variation in the residuals. *International Journal of Health Geographics*, 2007.

John G Ayisi, Anna H van't Hoog, Janet A Agaya, Walter Mchembere, Peter O Nyamthimba, Odylia Muhenje, and Barbara J Marston. Care seeking and attitudes towards treatment compliance by newly enrolled tuberculosis patients in the district treatment programme in rural western kenya: a qualitative study. *BMC public health*, 11(1):515, 2011.

R JB Baliddawa, Liefooghe, EM Kipruto, C Vermeire, and AO De Munynck. From their own perspective. a kenyan community's perception of tuberculosis. *Tropical medicine & international health*, 2(8):809–821, 2003.

Sudipto Banerjee. *Spatial Autoregressive Models*. 2009.

Sanjay Basu, David Stuckler, Asaf Bitton, and Stanton A Glantz. Projected effects of tobacco smoking on worldwide tuberculosis control: mathematical modelling analysis. *BMJ: British Medical Journal*, 343, 2011.

Bernardinelli, D Clayton, C Pascutto, C Montomoli, M Ghislandi, and L Songini, M. Bayesian analysis of spacetime variation in disease risk, 1995.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion) jr statist. *Soc. B*, 36:192–236, 1974.

Julian Besag and James Newell. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 143–155, 1991.

Nicky Best. Bayesian hierarchical modelling using winbugs. 2011.

Smith Brian. Bayesian output analysis program (boa) version 1.1 users manual. *Department of Biostatistics, College of Public Health, University of Iowa*, 2003.

James W Carey, Margaret J Oxtoby, Lien Pham Nguyen, Van Huynh, Mark Morgan, and Marva Jeffery. Tuberculosis beliefs among recent vietnamese refugees in new york state. *Public Health Reports*, 112(1):66, 1997.

Slide CDC Training. 70: directly observed therapy (dot). *Core Curriculum on Tuberculosis*, 2000.

David Clayton and John Kaldor. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, pages 671–681, 1987.

Peter Congdon. *Applied Bayesian hierarchical methods*. CRC Press, 2010.

Noel AC Cressie. *Statistics for Spatial Data, revised edition*, volume 928. Wiley, New York, 1993.

Christine SM Currie, Brian G Williams, Russell CH Cheng, and Christopher Dye. Tuberculosis epidemics driven by hiv: is prevention better than cure? *Aids*, 17(17):2501–2508, 2003.

Daniel and TM. The origins and precolonial epidemiology of tuberculosis in the americas: can we figure them out? unresolved issues. *The International Journal of Tuberculosis and Lung Disease*, 4(5):395–400, 2000.

Walter R David, Gilks, Sylvia Richardson, and Spiegelhalter. *Markov Chain Monte Carlo in practice: interdisciplinary statistics*, volume 2. Chapman & Hall/CRC, 1995.

Davis and AL. A historical perspective on tuberculosis and its control. *LUNG BIOLOGY IN HEALTH AND DISEASE*, 144:3–54, 2000.

Chaisson De Cock and RE KM. Will dots do it? a reappraisal of tuberculosis control in countries with high rates of hiv infection counterpoint. *The International Journal of Tuberculosis and Lung Disease*, 3(6):457–465, 1999.

Nyakeriga Emmanuel and 2Anakalo Shitandi. Spatial prevalence rate of tuberculosis as presenting to a rural hospital. *Journal of Applied Sciences Research*, 6(8):1008–1011, 2010.

Ezinna Enwereji. Views on tuberculosis among the igbo of nigeria. *Indigenous Knowledge and Development Monitor*, 7, 1999.

Paul Farmer. Social scientists and the new tuberculosis. *Social science & medicine*, 44(3): 347–358, 1997.

Paul Farmer, Simon Robin, St-L Ramilus, Jim Yong Kim, et al. Tuberculosis, poverty, and" compliance": lessons from rural haiti. In *Seminars in respiratory infections*, volume 6, page 254, 1991.

Maria Jacirema Ferreira Gonçalves, Antonio Carlos Ponce de Leon, and Maria Lúcia Fernandes Penna. A multilevel analysis of tuberculosisassociated factors. *Revista de Salud Pública*, 11(6):918–930, 2009.

José Figueroa-Munoz, Karen Palmer, MR Poz, Leopold Blanc, Karin Bergström, Mario Raviglione, et al. The health workforce crisis in tb control: a report from high-burden countries. *Hum Resour Health*, 3(2), 2005.

Janet Fleischman. Lessons from kenya for the global health initiative. 2011.

Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6): 721–741, 1984.

Jeff Gill. *Bayesian methods: A social and behavioral sciences approach*. CRC press, 2002.

Olivier Gimenez, Simon J Bonner, Ruth King, Richard A Parker, Stephen P Brooks, Lara E Jamieson, Vladimir Grosbois, Byron JT Morgan, and Len Thomas. Winbugs for population ecologists: Bayesian modeling using markov chain monte carlo methods. *Modeling demographic processes in marked populations*, pages 883–915, 2009.

Judith R Glynn, Kristin Kremer, Martien W Borgdorff, Mar Pujades Rodriguez, and Dick van Soolingen. Beijing/w genotype mycobacterium tuberculosis and drug resistance. *Emerging Infect. Dis*, 12(5):736–43, 2006.

John M Grange. The global burden of tuberculosis. *Porter, JDH and Grange, JG, Ed*, pages 1–33, 1999.

William A Hawley, Penelope A Phillips-Howard, FEIKO O TER KUILE, Dianne J Terlouw, John M Vulule, Maurice Ombok, Bernard L Nahlen, John E Gimnig, Simon K Kariuki, Margarette S Kolczak, et al. Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western kenya. *The American Journal of Tropical Medicine and Hygiene*, 68(4 suppl):121–127, 2003.

Parviez R Hosseini, André A Dhondt, and Andy P Dobson. Spatial spread of an emerging infectious disease: conjunctivitis in house finches. *Ecology*, 87(12):3037–3046, 2006.

Mollié Julian Besag, York. Bayesian image restoration, with two applications in spatial statistics. *Journal of Applied Statistics*, (5-6):21–59, 1991.

huong LM K, Lonnroth, PD Linh, and VK Diwan. Delay and discontinuity-a survey of tb patients search of a diagnosis in a diversified health care system. *The international journal of tuberculosis and lung disease*, 3(11):992–1000, 1999.

Leonhard Knorr-Held. Bayesian modelling of inseparable space-time variation in disease risk. 1999.

Leonhard Knorr-Held, Julian Besag, et al. Modelling risk from a disease in time and space. *Statistics in medicine*, 17(18):2045–2060, 1998.

Minjung Kyung and Sujit K Ghosh. Bayesian inference for directional conditionally autoregressive models. *Bayesian Analysis*, 4(4):675–706, 2009.

Stephen D Lawn and Robert Wilkinson. Extensively drug resistant tuberculosis. *Bmj*, 333(7568): 559–560, 2006.

Andrew B Lawson. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*, volume 20. Chapman & Hall/CRC, 2008.

Andrew B Lawson and AB Lawson. *Statistical methods in spatial epidemiology*. John Wiley, 2001.

Andrew B Lawson, William J Browne, and Carmen L Vidal Rodeiro. *Disease mapping with WinBUGS and MLwiN*, volume 11. Wiley, 2003.

Knut Lönnroth, Kenneth G Castro, Jeremiah Muhwa Chakaya, Lakhbir Singh Chauhan, Katherine Floyd, Philippe Glaziou, and Mario C Raviglione. Tuberculosis control and elimination 2010–50: cure, care, and social development. *The Lancet*, 375(9728):1814–1829, 2010.

Antonio López-Quılez and Facundo Munoz. Review of spatio-temporal models for disease mapping. 2009.

Leonardo Mariella and Marco Tarantino. Spatial temporal conditional auto-regressive model: A new autoregressive matrix. *Australian Journal of Statistics*, 39(3):223, 2010.

Elliot Marseille, G Kahn, Sharone Beatty, and Paul Perchal. Assessing the costs of multiple program approaches and service delivery modes for adult male circumcision in nyanza province, kenya. *New York: EngenderHealth*, 2011.

Elizabeth Marum, Miriam Taegtmeyer, and Kenneth Chebet. Scale-up of voluntary hiv counseling and testing in kenya. *JAMA: the journal of the American Medical Association*, 296(7):859–862, 2006.

T W Mazonde, Steen and TW Mazonde, GN. Ngaka ya setswana, ngaka ya sekgoa or both? health seeking behaviour in batswana with pulmonary tuberculosis. *Social science & medicine*, 48(2):163–172, 1999.

Harvey J Miller. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2):284–289, 2004.

Brian M Murphy, Denise, Benjamin H Singer, and Kirschner. On treatment of tuberculosis in heterogeneous populations. *Journal of Theoretical Biology*, 223(4):391–404, 2003.

JONATHAN D. Norton. Spatiotemporal bayesian hierarchical models, with application to birth outcomes. *THE FLORIDA STATE UNIVERSITY*, 2008.

Ioannis Ntzoufras. *Bayesian modeling using WinBUGS*, volume 698. Wiley, 2011.

Esa Nummelin. *General irreducible Markov chains and non-negative operators*, volume 83. Cambridge University Press, 2004.

World Health Organization et al. 2012 tuberculosis global facts. *Geneva, Switzerland*, 2011.

Katherine Ott. *Fevered lives: tuberculosis in American culture since 1870*. Harvard University Press, 1996.

Dirk Pfeiffer, Timothy Robinson, Mark Stevenson, Kim B Stevens, David J Rogers, and Archie CA Clements. *Spatial analysis in epidemiology*. Oxford University Press New York, 2008.

John Porter, Jessica A Ogden, Paul Pronyk, JDH Porter, JM Grange, et al. The way forward: an integrated approach to tuberculosis control. *Tuberculosis: an interdisciplinary perspective.*, pages 359–378, 1999.

Matthews AP Ramkissoon, Mwambi HG. Modelling hiv and mtb co-infection including combined treatment strategies. *PLoS ONE 7(11): e49492. doi:10.1371/journal.pone.0049492*, 2012.

Rindra Randremanana, Vincent Richard, Fanjasoa Rakotomanana, Philippe Sabatier, and Dominique Bicout. Bayesian mapping of pulmonary tuberculosis in antananarivo, madagascar. *BMC infectious diseases*, 10(1):21, 2010.

Mario C Raviglione. The new stop tb strategy and the global plan to stop tb, 2006-2015. *Bulletin of the World Health Organization*, 85(5):327–327, 2007.

Lee B Reichman and Earl S Hershfield. *Tuberculosis: a comprehensive international approach*, volume 144. Informa Healthcare, 2000.

Daiane Leite da Roza, Maria do Carmo Gullaci Caccia-Bava, Edson Zangiacomi Martinez, et al. Spatio-temporal patterns of tuberculosis incidence in ribeirão preto, state of são paulo, south-east brazil, and their relationship with social vulnerability: a bayesian analysis. *Revista da Sociedade Brasileira de Medicina Tropical*, 45(5):607–615, 2012.

Arthur J Rubel and Linda C Garro. Social and cultural factors in the successful control of tuberculosis. *Public health reports*, 107(6):626, 1992.

Frank Ryan. The forgotten plague: How the battle against tuberculosis was won-and lost author: Frank ryan, publisher: Back bay book. 1994.

María S Sánchez, James O Lloyd-Smith, Brian G Williams, Travis C Porco, Sadie J Ryan, Martien W Borgdorff, John Mansoer, Christopher Dye, Wayne M Getz, et al. Incongruent hiv and tuberculosis co-dynamics in kenya: Interacting epidemics monitor each other. *Epidemics*, 1(1): 14–20, 2009.

Birgit Schrödle and Leonhard Held. Spatio-temporal disease mapping using inla. *Environmetrics*, 22(6):725–734, 2011.

Schwarz. Establishment of a tuberculosis research data bank and information service. 1980.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

Guido Schwarzer, Gerd Antes, and Martin Schumacher. Inflation of type i error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Statistics in medicine*, 21(17):2465–2477, 2002.

Teeku Sinha and S Tiwari. Dots compliance by tuberculosis patients in district raipur (chhattisgarh). *Online Journal of Health and Allied Sciences*, 9(3), 2010.

Smith and . R. L. *Enviromental Statistics.* 2001.

David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

P. R. Srinivasan, Venkatesan and C. Dharuman. Bayesian conditional auto regressive model for mapping tuberculosis prevalence in india. *International Journal of Pharmaceutical Studies and Research*, 2012.

Daniel Thomas. *Captain of death: the story of tuberculosis*. University of Rochester Press, 1997.

Mukund Uplekar and Sheela Rangan. *Tackling TB: the search for solutions*. Foundation for Research in Community Health, 1996.

Norbert L Vecchiato. Sociocultural aspects of tuberculosis control in ethiopia. *Medical Anthropology Quarterly*, 11(2):183–201, 2008.

Best NG Wakefield, Elliot, DJ Briggs, et al. *Spatial epidemiology: methods and applications*. Oxford University Press, 2000.

NG Wakefield, Best and JC Waller, L. Bayesian approaches to disease mapping. *Spatial epidemiology: methods and applications*, pages 104–107, 2000.

Lance A Waller and Carol A Gotway. *Applied spatial statistics for public health data*, volume 368. Wiley-Interscience, 2004.

Lance A Waller, Bradley P Carlin, Hong Xia, and Alan E Gelfand. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92(438):607–617, 1997.

Yiyi Wang and Kara M Kockelman. A poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention*, 2013.

Helen Ward, Thierry E Mertens, and Carol Thomas. Health seeking behaviour and the control of sexually transmitted disease. *Health Policy and Planning*, 12(1):19–28, 1997.

Joseph HA Wilce, Shrestha-Kuwahara, JW Carey, R M Plank, and E Sumartojo. Tuberculosis research and control. *Encyclopedia of medical anthropology: health and illness in the worlds cultures. New York: Kuwer Academic/Plenum Publishers*, pages 528–542, 2004.

M Wilce, Shrestha-Kuwahara, N DeLuca, R, and Z Taylor. Factors associated with identifying tuberculosis contacts. *The International Journal of Tuberculosis and Lung Disease*, 7(Supplement 3):S510–S516, 2003.

David Wilkinson, L Gcabashe, and M Lurie. Traditional healers as tuberculosis treatment supervisors: precedent and potential planning and practice. *The International Journal of Tuberculosis and Lung Disease*, 3(9):838–842, 1999.