Bayesian Hierarchical Modelling with application in Spatial Epidemiology

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science

in

Mathematical Statistics

Department of Statistics

Rhodes University

by

Richard Southey

November 2017

Supervisor: Prof. L. Raubenheimer

Co-supervisor: Prof. S. E. Radloff

Declaration

I, the undersigned, declare that the work contained in this thesis is my own work, except for references specifically indicated in the text, and that I have not previously submitted it elsewhere for degree purposes.

_____ _____

R. R. Southey Date

Abstract

Disease mapping and spatial statistics have become an important part of modern day statistics and have increased in popularity as the methods and techniques have evolved. The application of disease mapping is not only confined to the analysis of diseases as other applications of disease mapping can be found in Econometric and financial disciplines. This thesis will consider two data sets. These are the Georgia oral cancer 2004 data set and the South African acute pericarditis 2014 data set. The Georgia data set will be used to assess the hyperprior sensitivity of the precision for the uncorrelated heterogeneity and correlated heterogeneity components in a convolution model. The correlated heterogeneity will be modelled by a conditional autoregressive prior distribution and the uncorrelated heterogeneity will be modelled with a zero mean Gaussian prior distribution. The sensitivity analysis will be performed using three models with conjugate, Jeffreys' and a fixed parameter prior for the hyperprior distribution of the precision for the uncorrelated heterogeneity component. A simulation study will be done to compare four prior distributions which will be the conjugate, Jeffreys', probability matching and divergence priors. The three models will be fitted in WinBUGS® using a Bayesian approach. The results of the three models will be in the form of disease maps, figures and tables. The results show that the hyperprior of the precision for the uncorrelated heterogeneity and correlated heterogeneity components are sensitive to changes and will result in different results depending on the specification of the hyperprior distribution of the precision for the two components in the model. The South African data set will be used to examine whether there is a difference between the proper conditional autoregressive prior and intrinsic conditional autoregressive prior for the correlated heterogeneity component in a convolution model. Two models will be fitted in WinBUGS® for this comparison. Both the hyperpriors of the precision for the uncorrelated heterogeneity and correlated heterogeneity components will be modelled using a Jeffreys' prior distribution. The results show that there is no significant difference between the results of the model with a proper conditional autoregressive prior and intrinsic conditional autoregressive prior for the South African data, although there are a few disadvantages of using a proper conditional autoregressive prior for the correlated heterogeneity which will be stated in the conclusion.

Keywords: Acute Pericarditis, Bayesian Statistics, Conditional Autoregressive Model, Disease Mapping, Oral Cancer, Standardised Mortality Ratio.

Contents

List of Figures

List of Tables

Abbreviations

List of Abbreviations

AIDS - Acquired Immune Deficiency Syndrome.
ASDR - Age-Specific Death Rate.
ANOVA - Analysis of Variance.
BGR - Brooks-Gelman-Rubin.
CAR - Conditional Autoregressive.
CDC - Centers for Disease Control and Prevention.
CDF - Cumulative Distribution Function.
CDR - Crude Death Rate.
CH - Correlated Heterogeneity.
CMR - Comparative Mortality Ratio.
DIC - Deviance Information Criterion.
DNA - Deoxyribonucleic Acid.
DSR - Directly Standardised Death Rate.
GEE - Generalised Estimating Equation.
GIS - Geographic Information System.
GLM - Generalised Linear Model.
GMRF - Gaussian Markov Random Field.
GOF - Goodness-of-Fit.
GP - Generalised Poisson.
HPA - Health Professions Act.
HPCSA - Health Professions Council of South Africa.
IAR - Indirectly Age-Adjusted Rate.
ICAR - Improper Conditional Autoregressive.
ICC - Intraclass Correlation Coefficient.
IID - Independent and Identically Distributed.
KL - Kullback-Leibler.
LMIC - Low-and Middle-Income Country.
LR - Likelihood Ratio.

MAPE - Mean Absolute Predictive Error.
MCMC - Markov Chain Monte Carlo.
MSPE - Mean Squared Predictive Error.
NCHS - National Center for Health Statistics.
NSAID - Nonsteroidal Anti-inflammatory Drug.
PCAR - Proper Conditional Autoregressive.
PDF - Probability Density Function.
PMP - Probability Matching Prior.
PP - Posterior Expected Exceedance Probability.
QR - Quantile Ratio.
Q-Q - Quantile-Quantile.
RR - Relative Risk.
SARS - Severe Acute Respiratory Syndrome.
SMR - Standardised Mortality Ratio.
UH - Uncorrelated Heterogeneity.
WHO - World Health Organization.

Notations

List of notations

y_i - Count of event in the i^{th} region.

μ_i - The average count in the i^{th} region.

I - Moran's Autocorrelation Coefficient.

ρ_0 - Pearson's Correlation Coefficient.

w_{ij} - Weight assigned to pair (y_i, y_j).

S_0 - Sum of all weights.

N - Total number of regions.

W - Matrix of weights with elements w_{ij}.

n - Number of age groups.

A_i - Grouping in the i^{th} region of the population of interest.

ζ_i - Age-group specific death rate in the i^{th} region.

N_{is} - Count in the i^{th} region of the standard population.

m_i - Count in the i^{th} region of the population of interest.

M - Matrix with conditional covariance proportional to $S_i|S_j$.

m_{ii} - Element of the matrix M.

N_{ps} - Total number of people in the standard population.

n_{is} - Middle of year population in the standard population.

Y_i - Count of event in the i^{th} region of the standard population.

e_i - Expected count in the i^{th} region.

θ_i - Estimated relative risk in the i^{th} region.

L - Likelihood.

l - Log likelihood.

$\ominus_i$ - Event rate per annum in the i^{th} region.

r_i - Cumulative events in the i^{th} region.

k_i - Number of person years at risk in the i^{th} region.

N_B - Number of iterations for the burn-in period.

n_T - Number of iterations.

n_A - Number of iterations to determine the DIC.

η_i - Link function for i^{th} observation.

v_i - Uncorrelated random effect at the i^{th} region.

ϕ - Precision of v_i.

S - Correlated heterogeneity.

ρ - Amount of spatial dependence.

ne_i - Number of neighbours of the i^{th} region.

r_{pi} - i^{th} Pearson residual.

r_{Bi} - i^{th} Bayesian residual.

r_{di} - i^{th} Deviance residual.

Acknowledgements

I would like to thank and acknowledge the work and support provided by Prof. Lizanne Raubenheimer. Her open door policy meant that there was always help and guidance when I needed it. This thesis would not have been possible without her advice, motivation, ideas and her problem-solving ability. When I was going around in circles, she always steered me in the right direction. I would also like to thank Prof. Sarah Radloff for her help and advice on generalised linear models and Prof. Andrew Lawson for his recommendation on the parameters of the hyperprior distributions. The financial assistance of the National Research Foundation (NRF) is hereby acknowledged, where my Master's studies have been funded by the NRF/SASA: Crisis in academic statistics bursary. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Chapter 1

Introduction

1.1 Overview

The evolution and discovery of new diseases is a continuous problem in the world today. The methodology and technology used to examine these diseases have experienced a rapid transformation in the past decade. Many different techniques are available to examine diseases, some of which are complex and expensive to perform. A relatively straightforward and fast way to analyse a disease is using disease mapping.

Disease mapping is a useful tool in the analysis and detection of patterns in diseases and involves mapping characteristics of a disease onto a region where the disease occurs. More recently, disease mapping has increased in popularity as the computational power of computers has improved along with the software used to produce disease maps. Evidence of the evolution of disease mapping is shown in the large amount of free and open source software available which can produce disease maps. The method is popular as it is a good way to visualise the characteristics of a disease. There are many different applications of disease mapping, some of which are more popular than others. The application which applies to this thesis is the analysis of the relative risk of contracting and dying from a disease. Two diseases are considered, oral cancer in Georgia in 2004 and acute pericarditis in South Africa in 2014.

A Bayesian approach will be used in this thesis. The Bayesian approach requires a prior distribution to be used along with the likelihood function to derive the posterior distribution, the conjugate prior and Jeffreys' prior will be used as the prior distributions and will be combined with the Poisson likelihood to derive the posterior distributions. The reasons for using the conjugate prior mainly stem from the use of WinBUGS® which requires a closed form prior to be specified. The use of the Poisson likelihood is justified in Section 2.8. There are many different methods to perform simulations. These methods include Rejection Sampling, Gibbs Sampling, Importance Sampling and many other numerical methods. Any choice of method will give a similar result. WinBUGS® automatically selects and performs this analysis.

1.2 Objectives and Contributions

This thesis will cover two objectives. The first objective is to examine the sensitivity of the hyperprior distribution of the precision of the uncorrelated heterogeneity and correlated heterogeneity. The first objective will be done using three prior distributions which are the Jeffreys' prior, conjugate prior and a fixed parameter prior. The fixed parameter is used to analyse the effect of not allowing estimation and flexibility in the model, as a prior distribution would typically allow. The three priors will be used to create three models on the Georgia oral cancer 2004 data set. The second objective will be to assess whether there is a difference between the improper conditional autoregressive (ICAR) and proper conditional autoregressive (PCAR) models. The correlated heterogeneity will be assigned the PCAR and ICAR models as a prior distribution, and the Jeffreys' prior will be used as the hyperprior for the precision of both the uncorrelated heterogeneity and correlated heterogeneity components.

The first contribution to the field of spatial epidemiology is that the hyperprior of the precision of the uncorrelated heterogeneity was sensitive to changes, although the Jeffreys' prior and conjugate prior produced comparable results. The prior with a fixed parameter and the other two priors had significantly different results for the Georgia oral cancer 2004 data set. The hyperprior of the precision for the correlated heterogeneity was more sensitive than the hyperprior of the precision of the uncorrelated heterogeneity and thus should be selected carefully. The second contribution to the field of spatial epidemiology is that there was no significant difference in the results for the South African acute pericarditis 2014 data set for the correlated heterogeneity with ICAR and PCAR prior distributions.

1.3 Thesis Outline

Chapter 2 provides the definitions and explanations of terms that are relevant to this thesis. A description and overview of disease mapping are provided in Section 2.2. In Chapter 2, definitions and equations used in this thesis are defined, it provides a concise but brief overview of epidemiology, the fundamentals of spatial epidemiology and a summary of the Poisson distribution. Other topics relevant to spatial analysis are also included in this chapter. Some of the fundamentals of Bayesian statistics are also defined in Chapter 2, and this definition includes the well-known Bayes' Theorem, different types of priors as well as the advantages and disadvantages of using these priors. An analysis to compare prior distributions is done in Chapter 2. This chapter also mentions different models which can be used in spatial epidemiology.

The data analysis and results are stated in Chapter 3. The chapter explains the limitations and challenges of the type of data used in spatial epidemiology. The results of the models are given in disease maps, and tables and conclusions are drawn. Summary statistics are also provided as well as interpretations.

Concluding remarks and findings are given in Chapter 4, possibilities for further investigations are

also stated in this chapter.

Appendix A contains the certificates used to collect the data and Appendix B contains the data. Further results are included in Appendix C. The RStudio ® and WinBUGS ® code is provided in Appendix D.

Chapter 2

Literature Review

2.1 Introduction into Epidemiology

Epidemiology is defined by the World Health Organization (2017) as any study of the causes and the distribution of diseases or health events and the use of such a study in the prevention, control and precautionary measures taken in disease control. Rothman (2012) summarises the definition of epidemiology as the "study of the occurrence of an illness."

The first known epidemiologist was John Graunt (1620-1674). His work included summarising data collected about the plague in England. This data had been collected for a number of years prior to the birth of Graunt. Graunt published research and summarised his results and findings, which was titled, according to Rothman (2012), "Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality." In Graunt's findings, the first estimate of the London population was done, and he wrote the first reports for time trends in which he adjusted for the population sizes. His findings were a breakthrough in epidemiology research. Other notable names mentioned in Rothman (2012) are Wade Hampton Frost (1880-1938), Janet Lane-Claypon (1877-1967), Florence Nightingale (1820-1910) and John Snow (1813-1858).

Centers for Disease Control and Prevention (CDC) (2012) says that a patient in a study is the community and all of the individuals in the study are put together collectively. The main characteristic of epidemiology research is that the data is collected for a disease that has a population at risk, according to Coggon et al. (2009). This population includes healthy and sick patients who are counted as if they had the study disease. The study requires a target population which measurements are not usually drawn from and only conclusions should be drawn from the target population. The measurements for any epidemiological investigation must be selected from a study population, which may be chosen from the target population. The public health problems associated with epidemiological studies are listed in Centers for Disease Control and Prevention (CDC) (2012) as:

- Environmental exposures

- Infectious diseases
- Injuries
- Noncontagious diseases
- Natural disasters
- Terrorism

The 21^{st} century saw a significant increase in the number of epidemiological studies done. The increase was due to improvements in computers, data collection and epidemiological methods and techniques. Studies previously had little attention from the media until the 1950's when some studies had a significant impact on various diseases worldwide, a few of these studies are named in Rothman et al. (2008) as:

- Avian Influenza
- Severe acute respiratory syndrome (SARS)
- Vaccination and autism
- Passive smoking and health
- Acquired immune deficiency syndrome (AIDS)

The increase in the number of studies has not been without controversy as the reliability and validity of methods used have often been questioned. This is most notable in the differences which arise between randomised studies in the Women's Health Initiative Randomised Controlled Trial done by the Writing Group for the Women's Health Initiative Investigators (2002) and unrandomised studies done by Stampfer & Colditz (1991). The goal of epidemiological studies is stated by MacMahon & Pugh (1970) as an examination of the relationships which may provide information into disease prevention.

2.2 Disease Mapping

Lawson et al. (2003) mention that the earliest application of disease mapping was done by Snow in 1854, where a map of addresses was put together for cholera victims relating to the proximity of the water supply. Recently, disease mapping has experienced substantial growth as it is a vital tool for disease prevention. Present-day disease mapping can incorporate geographic information systems as well as many statistical applications and procedures together. In disease mapping, confounders arise because of the nature of populations. Confounders are defined by Lesaffre & Lawson (2012) as known

explanatory variables which account for the influence of variables affecting the study population. The confounders are considered in a disease mapping sense via the introduction of random effects into the model. Other variables to be considered are deviation indices. These are variables which are related to a range of "poverty-related explanatory variables."

Disease mapping has also spread across many topics including ecological, geographical, health and spatial studies. Lawson et al. (2003) separate the main activities of disease mapping and lists them as disease mapping, disease clustering and ecological analysis.

For the disease mapping activity, which is mainly used in the health sector, an estimate of the "true relative risk" is found over a region or map. This application is mainly used for disease prevention and resource allocation. Disease clustering is used to determine where diseases are clustered most in a region. Lawson et al. (2003) state that this application is mainly used in environmental assessments. An ecological analysis is primarily used in epidemiological research to provide a "geographical distribution of disease in relation to explanatory covariates, usually at an aggregated spatial level."

There are two situations listed by Lawson et al. (2003) which are most used in the geographic spread of disease. These are: firstly, a study has a fixed period for which it takes place and secondly, the location of the observations of interest are recorded. This type of study is called a case-event analysis. Often the locations are street addresses or postal codes, and the privacy of information is a limiting factor in these types of studies because of the sensitive nature of the information which is collected. To address the limitations of the case-event analysis study, the tract count analysis study may be used. The tract count analysis study mentioned by Lawson et al. (2003) is based on the premise that the counts of the observations of interest within a small region are often recorded instead of using the locations of street addresses. The regions are arbitrary and may be used for census information.

For this thesis, we downloaded the map shape files which are freely available from a number of websites including the Global Administrative Areas available at (`http://gadm.org/country`). The shape files were then imported into Quantum Geographic Information System (QGIS)®(2016) available at (`http://www.qgis.org/en/site/`) and imported into WinBUGS® developed by Lunn et al. (2000). These maps are then used as the disease maps for the study and provide a graphical representation which simplifies the analysis of the diseases which are analysed. The disease maps used in this thesis are used to analyse the posterior expected exceedance probability (PP) and relative risk (RR).

Figure 2.1 is an example of a disease map of the number of observed deaths in South Africa by province in 2014. The data are from Statistics South Africa (2015).

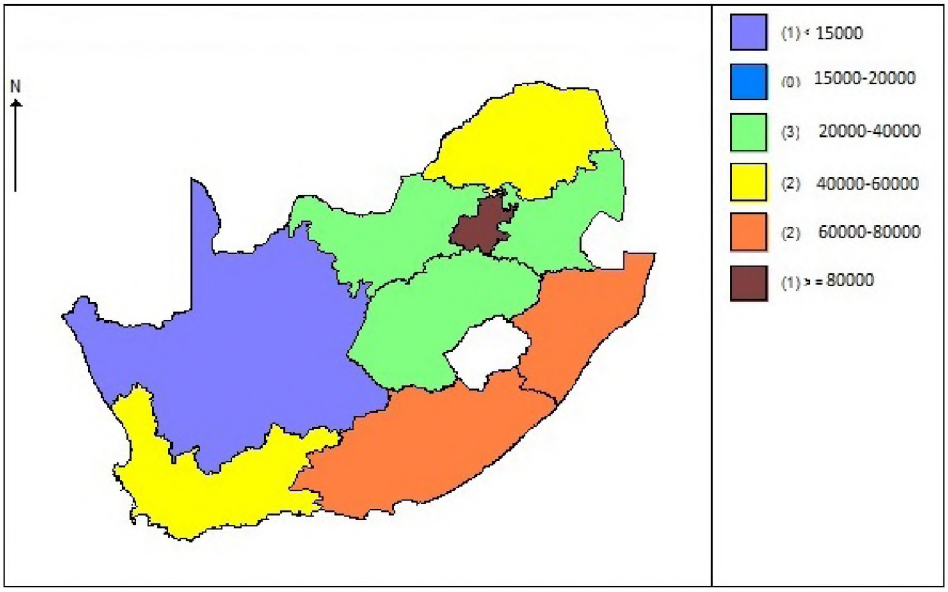


Figure 2.1: The observed number of deaths in South Africa by province in 2014.

Figure 2.2 is an example of further simplification of the observed number of deaths in South Africa in 2014 by municipal district, and the data are obtained from Statistics South Africa (2015).

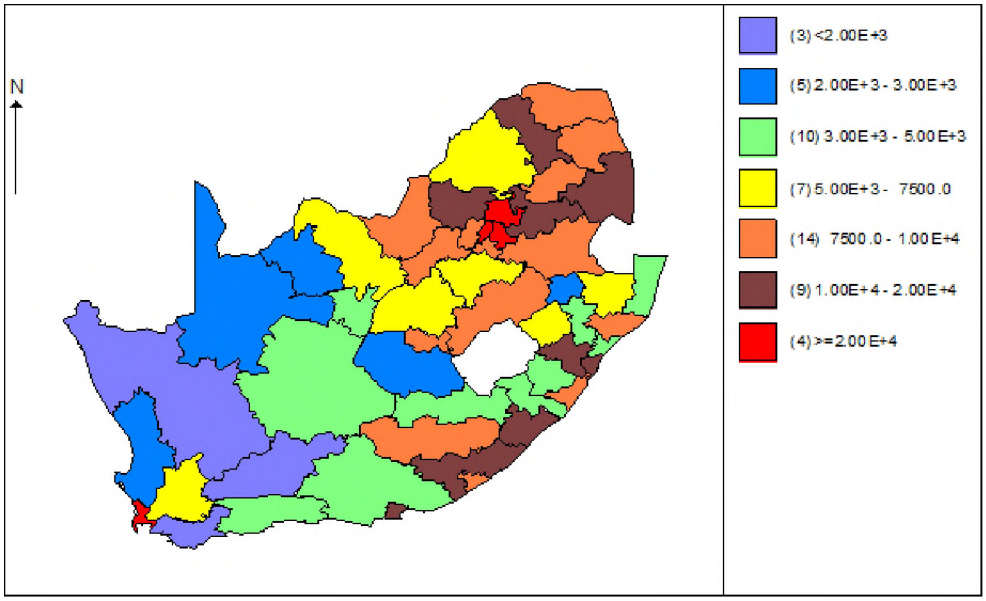


Figure 2.2: The observed number of deaths in South Africa by municipal district in 2014.

2.3 Disease Clustering

Clustering is defined by Lawson (2009) as "any spatially-bounded region of significantly elevated (reduced) risk." This relates to the correlated heterogeneity in relative risk models. This type of

clustering is used when the relative risk is similar in neighbouring regions. Clustering is used to determine whether there is a relationship between a region and the layout of a disease. Tango (2010) refers to this relationship as examining the distribution of a disease to determine whether the disease is distributed at random or whether clusters develop within a time period or in a region, provided confounding variables have been accounted for in the study.

Within the topic of clustering, there are different types of clustering, namely: global clustering and hot-spot clustering. Global clustering is referred to by Lawson (2009) to occur when a region has a peak in risk. Hot-spot clustering relates to any region that can be clustered and does not have the assumption of a neighbourhood criterion. Hot-spot clustering is often used in epidemiology when a disease is relatively new and the behaviour of the disease is not well understood. Other classifications are made by Tango (2010) who separates clustering into three main groups. These groups are temporal clustering, spatial clustering and space-time clustering. Temporal clustering is used to determine whether diseases occur in close proximity within a specific time period. Spatial clustering is used to determine whether diseases occur in close proximity to specific regions and space-time clustering is a combination of both spatial and temporal clustering.

Clustering may be used when there is some sort of grouping within the data set. The groupings are defined before the study and are known to the researcher. Sometimes it may be useful to define clusters in a residual term in the data. Lawson (2009) gives an example of this by letting y_i denote the count of deaths in the i^{th} group of a region. Let the average count be denoted by μ_i and the model be denoted by: $log(\mu_i) = \alpha_i + \epsilon_i$, where ϵ_i contains the residual terms. Then model α_i to contain all of the non-residual terms such that ϵ_i must contain all of the residual clustering terms, ϵ_i can consist of clustered and unclustered terms. A pure noise term can be introduced into α_i to isolate ϵ_i as the cluster term.

In this thesis, possible disease clusters are identified using the posterior expected exceedance probability.

Lawson (2014) states that clustering is used to:

- Find the cause of a disease.
- Detect disease cluster alarms.
- Assist in Public Health applications.

2.4 Spatial Correlation

2.4.1 Overview

Lesaffre & Lawson (2012) define spatial correlation as a "fundamental feature of Geo-referenced data" which comes about from a local human population which "varies in spatial density and in susceptibi-

lity" to the disease which is being investigated. Douglas et al. (2000) state that many studies do not account for spatial correlation (autocorrelation) in the interpretation of their results because: (1) the assumption of independent observations is often necessary for many experiments, (2) an overlap in the data is created and affects the data collected at each data point and this results in the sample size being reduced, (3) hypothesis testing is affected and according to Douglas et al. (2000) if the spatial correlation exists in the data, the probability of rejecting the null hypothesis may be higher than if there were no spatial correlation, which would lead to false results and (4) the correlation between data points also makes it difficult to determine a probability distribution of the test statistic as the assumption of independence is broken. Researchers often have to use an "approximate distribution." Lawson (2009) states that: (a) spatial correlation is geographical and comes about because locations that share the same boundaries may have related values of outcome variables whereas locations that are not in close proximity to other locations may have values of outcome variables which differ to other locations, (b) autocorrelation affects the formulation of the structure of the likelihood for regions that contain correlation and (c) an individual contribution to the likelihood may be assumed independent of other contributions and this allows the likelihood to be derived as the product of probabilities. The spatial distribution of a control disease may be used according to Lesaffre & Lawson (2012) for case event data because expected rates are unavailable and the spatial variation in the disease of interest is compared to the spatial variation of the control disease. The choice of control disease may be controversial. The control disease must be chosen with age-sex structures analogous to the disease of interest according to Lesaffre & Lawson (2012).

Besag et al. (1991) advise that the observed events can be considered as independent in the case where the disease of interest is deemed to be rare and noncontagious. Considering that there were 16 deaths from acute pericarditis in a population size of 54001953 in South Africa in 2014 and a crude death rate (CDR) of 0.000296 deaths per 1000 people. Then it is reasonable to assume that the event of death as a result of acute pericarditis is a rare event although the disease is frequently diagnosed. The nature and characteristics of acute pericarditis show that the disease is also noncontagious. Based on the information from Besag et al. (1991) it is, therefore, reasonable to assume independence for the observed events of acute pericarditis.

The nature and characteristics of oral cancer also show that the disease is noncontagious. For the Georgia oral cancer data in 2004, there are 218 deaths in total in a population size of 8769252. An examination of the CDR reveals that there are 0.02486 deaths per 1000 people in Georgia and therefore it is fair to assume that death from oral cancer is a rare event in Georgia and independence of observed events may be assumed. The technology and methods used to treat oral cancer are improving rapidly which contributes to the low number of deaths.

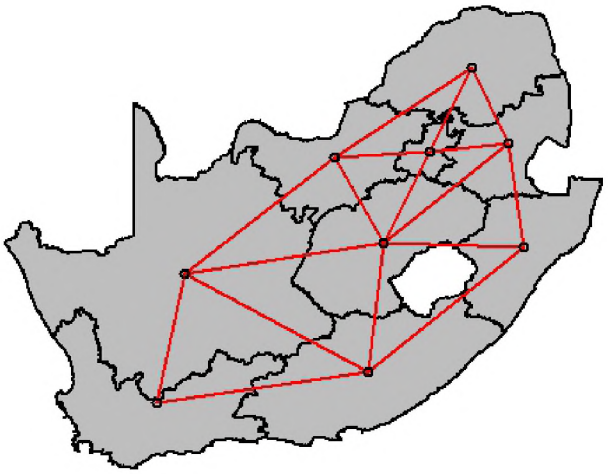


Figure 2.3: The neighbouring regions in South Africa.

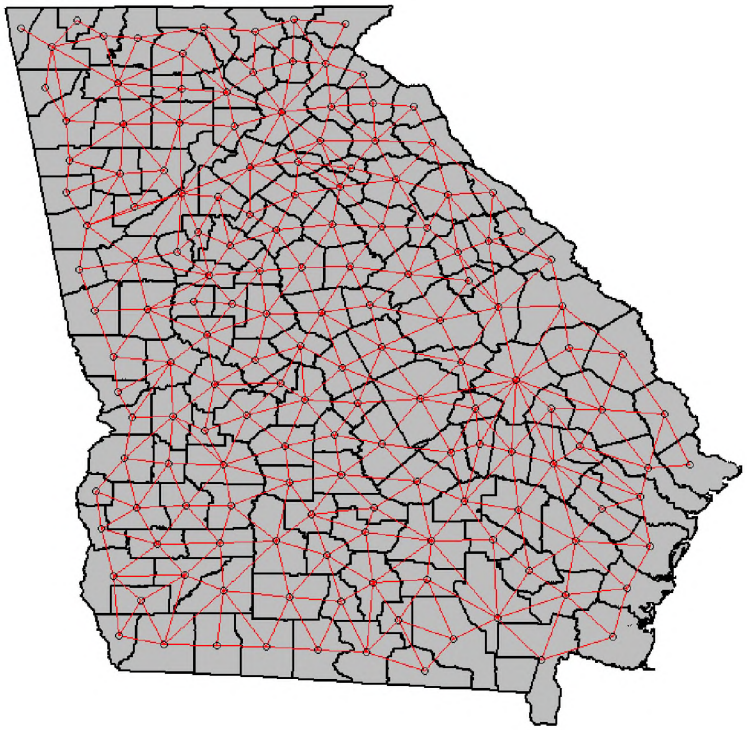


Figure 2.4: The neighbouring regions in Georgia.

Figures 2.3 and 2.4 show the neighbouring regions in South Africa and Georgia, respectively, and show the regions which share borders. These figures are produced using the packages *maptools*, written by Bivand & Lewin-Koh (2013), and *spdep*, written by Bivand & Piras (2015), in RStudio®. The code is based on the method by Hijmans (2016). The shape files for both South Africa and Georgia were downloaded from the Global Administrative Areas website available at (`http://gadm.org/country`). The shape file for both South Africa and Georgia are imported into RStudio® using the

$readShapeSpatial$ function from the $maptools$ package. This function converts the shape file into a spatial data frame object which can be read and edited in RStudio®. The regions of interest, provinces for South Africa and counties for Georgia, are then extracted from the spatial data frame object. This is easily done as it requires only matrix notation. The next step is to extract the coordinates for each of the regions. The $coordinates$ function in the $spdep$ package is used to extract this information. The function returns a matrix containing the x and y coordinates for the centres of each region. This is used for the nodes that are connected by the red lines in Figures 2.3 and 2.4 . The neighbours of each region are now required to complete the information needed for the figures. This is done by the $poly2nb$ function, available from the $spdep$ package, which creates a neighbours list object. The figures can now be plotted by plotting the shape file and then adding the centre of each region accompanied by the lines which connect them, which are obtained from the neighbours list object.

2.4.2 Moran's Autocorrelation Coefficient

The Moran's Autocorrelation Coefficient, developed by Patrick Alfred Pierce Moran (1950) in the paper named "Notes on Continuous Stochastic Phenomena", can be thought to be an application of the "Pearson product-moment correlation coefficient." The Pearson's correlation coefficient is well-known and given as:

$$\rho_0 = \frac{\sum_{i=1}^{n}\Big(x_i-\bar{x}\Big)\Big(y_i-\bar{y}\Big)}{\left[\sum_{i=1}^{n}\Big(x_i-\bar{x}\Big)^2\sum_{i=1}^{n}\Big(y_i-\bar{y}\Big)^2\right]^{\frac{1}{2}}}. \tag{2.1}$$

Equation 2.1 measures the strength of the linear relationship between two variables, x and y. The difference between Moran's Autocorrelation Coefficient, denoted by I, and the "Pearson product-moment correlation coefficient", ρ_0, is that Pearson's correlation coefficient measures the linear relationship between two independent variables, x and y, whereas Moran's Autocorrelation Coefficient measures the relationship between y_i and y_j where $i \neq j$ and thus in the "Pearson product-moment correlation coefficient" there will be no linear relationship between y_i and y_j when $i \neq j$ (Paradis, 2017). When observations are measured in close proximity to one another, it is to be expected that the relationship between these observations will be higher than observations that are further apart, to account for this Paradis (2017) suggests that weights, denoted by w_{ij}, must be assigned to each pair of observations. These weights are values of 0 or 1, with a weight of 1 given to pairs who are in close proximity and a value of 0 assigned to pairs who are in further proximity. The Moran's Autocorrelation Coefficient is well-known and given by:

$$I = \frac{n}{S_0} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\Big(y_i - \bar{y}\Big)\Big(y_j - \bar{y}\Big)}{\sum_{i=1}^{n}\Big(y_i - \bar{y}\Big)^2},$$

where the weight assigned to the pair $\Big(y_i,\ y_j\Big)$, where $i \neq j$, is denoted by w_{ij} and S_0 is the sum of all weights which are measured.

An adjustment can also be made for observations that are in close proximity to one another. This adjustment involves scaling the distance between all the pairs.

A model should be assessed to check if there is a residual spatial structure in the data after the model has been fitted. A model which fits the data adequately should have little spatial correlation in the residual of the data after the model has been fitted, as stated by Lawson (2009). In the case of a simulation, an estimate of the spatial structure may be obtained and averaged over the posterior sample. The Moran's Autocorrelation Coefficient is a popular choice for the assessment of the spatial structure. The Moran's Autocorrelation Coefficient is defined by Lawson (2009) for a simulation in terms of quadratic forms as:

$$I = \frac{\boldsymbol{res}' W \boldsymbol{res}}{\boldsymbol{res}' \boldsymbol{res}}, \tag{2.2}$$

where $\boldsymbol{res}$ is a vector of residuals containing the N residuals from each region and W is an adjacency matrix with 0 or 1 elements for the regions with elements w_{ij}.

2.5 Methods for Comparison of Populations

2.5.1 Overview

The mortality rate often needs to be compared over different geographical regions. These kinds of studies are most used by Medical Health Professionals. There are a number of methods which are used to compare mortality rates over different regions, one of the most commonly used methods is, according to Ahmad et al. (2001), (1) the age-specific rate over time and (2) the CDR, which is the total number of deaths per 1000 people in a year as defined by Sankoh et al. (2014) and is a function of the underlying age structure of the population. These comparisons, however, are inappropriate when a comparison is made between two populations with different age structures. Every geographic region has a different age structure, although regions which are in close proximity to each other may have very similar age structures with minor differences, the regions which are not in close proximity have notable differences in the underlying age structure. Waller & Gotway (2004) state that the age-specific rates of two or more regions may differ because of only the difference of the underlying age structure in those regions, instead of differentiating due to the "age-specific risk" of the disease.

Thus, as Julious et al. (2001) explain, any analysis of the crude mortality rate for these regions are affected by these confounding differences in age structures of the study populations. The differences in age structure are most notable between low-and middle-income countries (LMICs) and high-income countries, according to Sankoh et al. (2014). For these studies to be completed, the effect of the age structure of each of the study populations needs to be minimised. This minimisation is defined by Julious et al. (2001) as age standardisation.

Age standardisation allows population mortality rates to be compared irrespective of the underlying population structure and according to Waller & Gotway (2004), standardisation provides a method to remove the effect of age and enables a comparison of populations in different regions. Each population in the study region is divided into age and sex subsets. Julious et al. (2001) state that standardisation in the populations of interest is calculated by determining a weighted average of the subset specific mortality rate. The method of standardisation introduces a standard population. Examples of a standard population which are commonly used in research are named in Waller & Gotway (2004) as "superpopulation, containing the study population" and the "total subpopulation." The weights in standardisation show the relationship between the age distributions of the population of interest and the standard population.

Ahmad et al. (2001) name some of the methods for age standardisation as: (1) direct and indirect standardisation, (2) the geometric mean, (3) equivalent average death rates, (4) life table rates, (5) cumulative death rates and (6) the comparative mortality index. The most common techniques of standardisation are direct and indirect standardisation as stated by Julious et al. (2001). Standardisation is defined by Keiding (1987) as a method used when data are divided into n age groups and a standard population has $n_1, n_2, \ldots, n_n$ observations in each group. The $n_1, n_2, \ldots, n_n$ observations along with "age-group specific death rates" denoted by $\zeta_1, \zeta_2, \ldots, \zeta_n$ are contrasted with a population of interest with grouping denoted by $A_1, A_2, \ldots, A_n$ (Keiding, 1987). The expected number of deaths are then calculated by $\sum_{i=1}^{n} n_i \zeta_i$ and can be compared with the observed number of deaths in the population of interest. This comparison is the standardised mortality ratio (SMR). Fleiss et al. (1981) provide reasons for standardisation as:

- A single rate is easier to interpret and use than many rates in the same population.
- When the size of a region is small, then any rates calculated based on observations in that region may be unreliable for obtaining results over many regions.
- Some subset of a population which the study is to be conducted on may not allow for age-specific rates to be calculated.

2.5.2 Direct Standardisation

2.5.2.1 Direct Standardisation History

Direct standardisation, according to Ahmad et al. (2001), was developed because crude rates were deemed inappropriate in studies when the underlying age structure of a region was different to another region. There was thus the need to develop a measure which was independent of the age structure. Sir Edwin Chadwick made use of "the mean at death" as a measure to solve this problem in London, England. This measure was later shown by Neison (1844) to be dependent on the underlying age structures of the populations because the mortality increased with age and thus "the mean at death" measure was also inappropriate for such studies. Neison (1844) then developed the method of direct standardisation by making a comparison of the mean age with the "crude mean age at death". Neison (1844) also was the first person to use standard populations and indirect standardisation.

The first report of Neison's direct standardisation method was, according to Ahmad et al. (2001), in the Registrar General's report in 1883. The report made use of data in the 1881 population census of England and used the population of Wales as the standard population. It was later determined that one standard population needed to be used for the studies as in earlier studies a new standard population had to be calculated each time a new study was conducted. Studies then made use of a population census as a standard population. This was done in 1901.

2.5.2.2 Direct Standardisation Method

Direct standardisation is used by Carneiro et al. (2011) when the "group-specific" outcome rates of the population of interest are known. The calculation of which results in a standardised death rate which is a weighted average of the age-specific rates. This accounts for the different age structures of the populations of interest as stated by Ahmad et al. (2001). Waller & Gotway (2004) describe the problem which direct standardisation solves as "how many cases would we observe in the standard population if the observed age-specific rates of the disease applied?". Direct standardisation results in a rate that shows the number of deaths that we expect if the populations of interest had the same underlying age structure. The directly standardised death rate (DSR) is calculated by Julious et al. (2001) for the population of interest as:

$$DSR = \sum_{i=1}^{n} \frac{N_{is}}{N_{ps}} \frac{y_i}{m_i}, \tag{2.3}$$

where N_{is} is the count of people in the i^{th} group of the standard population, y_i is the count of deaths in the i^{th} group of the population of interest, m_i is the count of people in the i^{th} group of the population of interest, N_{ps} is the total count of people in the standard population and n is the number of groups. When comparing two populations which have similar underlying age groups, then the selection of the

standard population does not have a significant impact on age-groups or time periods. This is not usually the case in reality, thus the DSR is generally affected by the choice of the standard population, since each region or population may have small differences in the underlying age structure than another region or population. Ahmad et al. (2001) compare two populations, denoted by X and Y, by calculating each DSR as:

$$DSR_Y = \sum_{i=1}^{n} \zeta_{iY} \left[\frac{n_{is}}{\sum_{i=1}^{n} n_{is}} \right]$$

and

$$DSR_X = \sum_{i=1}^{n} \zeta_{iX} \left[\frac{n_{is}}{\sum_{i=1}^{n} n_{is}} \right],$$

where n_{is} is the population in the middle of the year in the standard population of the i^{th} age group, ζ_{iX} and ζ_{iY} are the death rates in each population for the i^{th} age group and n is the number of groups. A comparative mortality ratio (CMR) is calculated by the ratio of DSR_Y and DSR_X. CMR is calculated in Julious et al. (2001) as:

$$CMR = \frac{\sum_{i=1}^{n} N_{is} \frac{y_i}{m_i}}{\sum_{i=1}^{n} N_{is} \frac{Y_i}{N_{is}}}, \tag{2.4}$$

where N_{is}, y_i and m_i are given in Equation 2.3 and Y_i is the count of deaths in the i^{th} group of the standard population.

The confidence interval for CMR is calculated by Julious et al. (2001) as:

$$\left(\frac{CMR}{exp\left[\frac{Z_{\frac{\alpha}{2}} \times SE(CMR)}{CMR}\right]};\ CMR \times exp\left[\frac{Z_{\frac{\alpha}{2}} \times SE(CMR)}{CMR}\right] \right),$$

where $SE(CMR)$ is the standard error of the CMR and is calculated by:

$$SE(CMR) = \sqrt{\left(\sum_{i=1}^{n} N_{is}^2 \right) \frac{y_i}{m_i^2}},$$

where N_{is}, y_i and m_i are given in Equation 2.3. The log of the standard error of CMR is used by Julious et al. (2001) and is given by:

$$log\Big(SE(CMR)\Big) = \frac{SE(CMR)}{CMR}.$$

Waller & Gotway (2004) use direct standardisation when the following data are available:

- Age specific rates from the population of interest.

- The number of people who may experience an event in the standard population.
- The total number of people who observed an event in the standard population.

Figure 2.5 provides the idea behind direct standardisation and is found in Naing (2000).

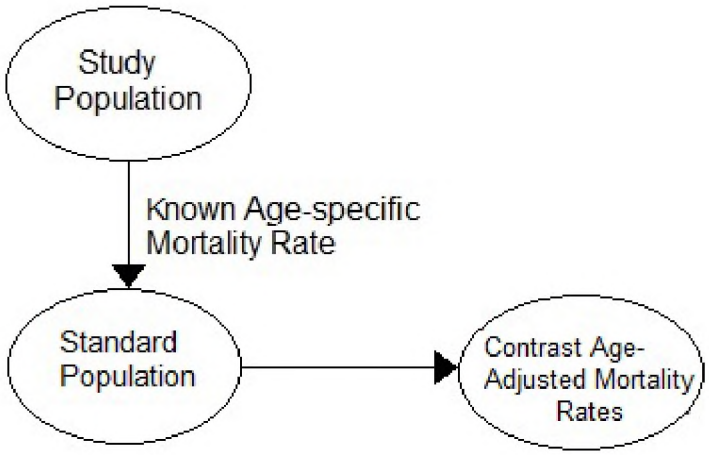


Figure 2.5: The direct standardisation method.

2.5.3 Indirect Standardisation

2.5.3.1 History of Indirect Standardisation

Keiding & Clayton (2014) state that indirect standardisation was originally proposed by Neison (1844). Neison later published a survey on "the rate of mortality among persons of intemperate habits" in 1851 which applied indirect standardisation. Another application of indirect standardisation was by Farr (1859) in which "age-specific death rates for 1849-1853" were used. H. Westergaard, a Danish economist and statistician, used indirect standardisation in his research named "Die Methode der Erwartungsmässig Gestorbenen" which translates to "The Method of Expected Deaths" as explained by Keiding & Clayton (2014). Westergaard also showed the importance of introducing confounders into a spatial model. Westergaard later introduced a method to calculate the standard error of the expected number of deaths. This expected number of deaths is important because it may be used to reduce the effect of age on the model and the standard error of the expected number of deaths may be used to measure the effectiveness of this.

2.5.3.2 Indirect Standardisation Method

Indirect standardisation is an approximation to direct standardisation and is used when the data is not available for direct standardisation (Curtin & Klein, 1995). Indirect standardisation is described by Waller & Gotway (2004) to answer the question of "what would the expected number of cases be in

the study population if people in the study population contracted the disease at the same rate as people in the standard population?". When a rare disease occurs in a population, the age-specific rate can be unstable, especially when the number of people within each region is small. The addition of one case to the data has a notable effect on the age-specific rate of that region and it is thus preferable to use indirect standardisation. Confidentiality and privacy of health data limit the use of direct standardisation and hence indirect standardisation becomes preferable as stated by Waller & Gotway (2004).

Indirect standardisation estimates the expected number of deaths in the population of interest as if the age-specific of a standard population had been applied (Carneiro et al., 2011). It is used when there are no specific rates for the population of interest but the total number of deaths and the population structure are both known. Indirect standardisation uses a standard age-specific death rate (ASDR) which is applied to the population of interest to calculate the expected counts. This assumes that the standard ASDR applies to the population.

Lesaffre & Lawson (2012) provide a method to account for autocorrelation and obtained expected rates for the disease which is being investigated based on the "age-sex structure of the local population." Estimates of the "local relative risks" may also be calculated from the ratio of observed to expected counts and are called the standardised mortality ratios (SMRs). Lawson et al. (2003) calculate SMRs by using the formula $\theta = \frac{y_i}{e_i}$ where y_i is the observed count in the i^{th} region, e_i is the expected count in the same region and θ is an estimate of the relative risk in that region. The SMR, in Equation 2.5, is derived using the likelihood by:

$$\begin{aligned} L\Big(e_i\theta_i|y_i\Big) &= \prod_{i=1}^{N} \frac{\Big(e_i\theta_i\Big)^{y_i} exp\Big(-e_i\theta_i\Big)}{y_i!} \\ l\Big(e_i\theta_i|y_i\Big) &= \sum_{i=1}^{N} y_i\Big(ln(e_i\theta_i)\Big) - \sum_{i=1}^{N}\Big(e_i\theta_i\Big) - \sum_{i=1}^{N} ln\Big(y_i!\Big), \end{aligned}$$

and differentiating this with respect to $e_i\theta_i$ and equating to zero:

$$\begin{aligned} \frac{\partial l\Big(e_i\theta_i|y_i\Big)}{\partial e_i\theta_i} &= \frac{\partial \sum_{i=1}^{N} y_i\Big(ln(e_i\theta_i)\Big) - \sum_{i=1}^{N}\Big(e_i\theta_i\Big) - \sum_{i=1}^{N} ln\Big(y_i!\Big)}{\partial e_i\theta_i} \\ &= \frac{y_i}{e_i\theta_i} - 1. \end{aligned}$$

So

$$\begin{aligned} 1 &= \frac{y_i}{e_i\theta_i} \\ y_i &= e_i\theta_i \\ \theta_i &= \frac{y_i}{e_i}. \end{aligned}$$

The SMR is calculated in Julious et al. (2001) as:

$$SMR = \frac{\sum_{i=1}^{N} m_i \frac{y_i}{m_i}}{\sum_{i=1}^{N} m_i \frac{Y_i}{N_{is}}}, \tag{2.5}$$

where $N_{is},\ y_i$ and m_i are given in Equation 2.3 and Y_i is given in Equation 2.4.

There are a number of disadvantages of using SMRs because a small change in the expected value causes a substantial change in the SMRs and when the expected value is estimated to be close to zero, a large SMR value is calculated. We conclude that the SMR is therefore unstable. The variance is also affected by small changes to the expected counts since the variance of the SMR is proportional to $\frac{1}{e_i}$. The SMR is zero when the observed count is zero but the expected value e_i may vary significantly. The SMR may also be biased depending on the type of information used in the model. The bias arises especially in models where the data are made up of people who have not been exposed to a disease together with those who have been exposed to the disease. This bias is investigated in Jones & Swerdlow (1998). Methods to improve the estimates of SMRs are covered in Lawson et al. (2003). The most common methods involve smoothing the SMR which requires adding extra components to the model. This smoothing is done by assigning a prior distribution to the model as stated by Lawson (2003). Conclusions based on the SMR will thus not be drawn for this thesis and conclusions will be based on the relative risk.

Waller & Gotway (2004) propose that indirect standardisation requires the following data:

- Age-specific rates for the standard population.
- The number of people who may experience an event in the population of interest.
- The total number of people who experienced an event in the population of interest.

In this thesis, the indirect standardisation method is used as there are no specific rates for the population of interest but the total number of deaths and the population structure are both known for the South African and Georgia data.

Figure 2.6 provides the idea behind indirect standardisation and is found in Naing (2000).

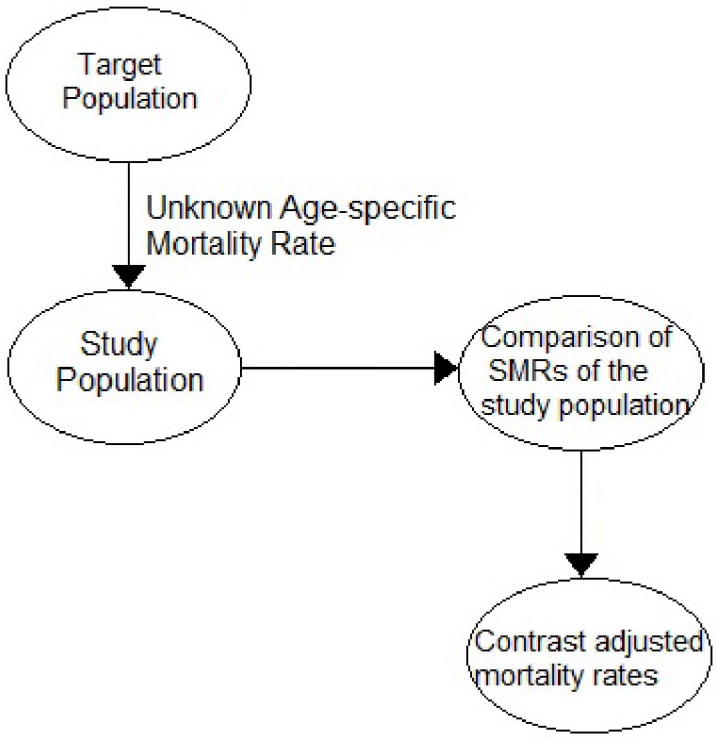


Figure 2.6: The indirect standardisation method.

2.5.4 Indirect Standardisation versus Direct Standardisation

Julious et al. (2001) state that SMR is not appropriate for comparing populations as the regions have different denominators. When comparing Equation 2.5 with Equation 2.4, one may note that SMR is calculated based on the population of interest and the CMR is calculated based on the standard population. The calculation of SMR depends on the weights of age and gender groups hence the SMR is dependent upon the population of interest. Thus, for different regions, the weights of the age and gender groups will be different and thus affect the SMR as explained by Julious et al. (2001). In contrast, the CMR will have the same weights, as the CMRs are calculated from the same standard population. This means that each SMR is standardised to a different population.

An important advantage of indirect standardisation is that the method does not require the age groups of the observations to be given as proposed by Keiding & Clayton (2014). When a comparison is made between indirect and direct standardisation, Pickle & White (1995) suggest expressing the equation for indirect standardisation in terms of the indirectly age-adjusted rate (IAR) at the i^{th} region as:

$$\begin{aligned} IAR_i &= SMR_i \begin{pmatrix} crude\ rate\ in \\ standard\ population \end{pmatrix} \\ &= \frac{\left(\frac{\sum y_i}{m_i}\right)\left(\frac{m_i}{\sum m_i}\right)\sum\left(\frac{Y_i}{N_{is}}\right)N_{is}}{\left(\frac{\sum Y_i}{N_{is}}\right)\left(\frac{m_i}{\sum m_i}\right)\sum N_{is}}. \end{aligned} \tag{2.6}$$

The first part of Equation 2.6 incorporates the proportion of weighted averages in which crude proportions of the population of interest and the standard population are weighted by the underlying age-distribution of the population of interest. The second part of the equation is a weighted average of the crude rate of the standard population where the weights are calculated from the age-distribution of the standard population. This shows that for indirect standardisation the weights for each population are different whereas in direct standardisation, weighting is made to be dependent upon the standard population. Indirect standardisation is therefore dependent on the underlying age distribution of the population of interest and may pose problems when a comparison is required to analyse two different populations in two different regions as explained by Waller & Gotway (2004).

There are certain and strict conditions where indirect and direct standardisation give similar results, these conditions are stated in Pickle & White (1995) as:

- When the age-specific rates in the population of interest are the same as the age-specific rate in the standard population.
- When the age-specific rates in the population of interest are determined to be proportional to the age-specific rate in the standard population.
- When the population distribution of the population of interest can be expressed as proportional to the population distribution of the standard population.

Julious et al. (2001) state that SMRs should only be used to compare regions or populations with the same age structures.

2.5.5 The Standard Population

2.5.5.1 History of the Standard Population

In Section 2.5.2.1, the need of one international standard population was explained. Historically, this lead to the 1901 population census in England being used as the standard population for both England and Wales.

Ahmad et al. (2001) state that the United States also used the 1901 England standard population in their studies. This was used until 1940 when it was proposed that the 1901 standard population was

vastly different to the population in the United States at the time. The United States then used its 1940 population census as its standard population. The main problem with the standard population was when to change it to a new standard population as the development of a new standard population was very time consuming and it was difficult to determine when the standard population still reflected the current population. A new standard population was later introduced as the 1940 standard population was deemed to not fully represent the current population at the time. This standard population was based on the year, 2000, population estimates. The new standard population was introduced by Rosenberg & Anderson (1998) in a report for the National Center for Health Statistics (NCHS).

There was a need for a standard population which could be applied to several countries, according to Ahmad et al. (2001). The idea of the international standard population was first proposed by Ogle (1892) in his article "Proposal for the Establishment and International Use of a Standard Population." Ogle (1892) used the population estimates of seven different European countries. His standard, however, was never widely used. Many more standard populations have been developed but never adopted because of the debate whether one standard population is favourable over another standard population, as explained by Ahmad et al. (2001). The solution to this debate was to create three standard populations with each standard population having a different purpose. The first standard population has a higher weighting in the lower age groups which is intended to be used for African countries and was developed by Davies et al. (1962). The second standard population was based on estimates from the Scandinavian population. This standard population had a higher weighting in the older age groups and was designed to be used in Western Europe. It was given the name "European" standard population and proposed by Doll & Cook (1967), according to Ahmad et al. (2001). The third standard population, named the "World" standard, was based on the populations of forty-six countries and was developed by Segi (1960). The "World" and "European" standards are used by the World Health Organization (WHO) in the calculation of age-standardised death rates. Table 2.1 shows the trend in standard population from $1970 - 1995$ and can be found in Ahmad et al. (2001):

Table 2.1: The changes in standard populations of US males (1970-1995) in age-adjusted circulatory disease mortality rates per 100 000 people.

Standard Population	**1970**	**1975**	**1980**	**1985**	**1990**	**1995**	**Change in percentage from 1970 – 1995**
Segi	459.5	399.0	350.3	305.8	256.8	232.3	−49.4
WHO World Standard	550.9	482.2	426.7	373.7	315.0	285.4	−48.2
Scandinavian	720.1	630.4	557.8	488.4	411.6	372.4	−48.3

2.5.5.2 Derivation of the Standard Population

Ahmad et al. (2001) introduce the idea that the standard population should reflect the average age-structure over 25-30 years of the population(s) of interest. The standard populations are therefore only re-estimated every 25-30 years to incorporate this average. The effect of historical events such

as war and plagues are removed by using the average world population. The United Nations Population Division calculates the population age-structures by age and sex for each country biannually. The average world population structure is based upon estimates of the populations of each of these countries. Estimation is done every 5 years from 1950 and estimated up until 2025.

2.5.5.3 The Choice of Standard Population

The choice of the standard population is important because different regions have different underlying age structures. The use of an inappropriate standard population results in a higher or lower weighting on age groups in the calculation of the ratios and hence leads to false results. For example, LMICs have a larger percentage of the population under the age of fifteen than higher income countries.

The standard population used for comparing two populations must be the same. Ahmad et al. (2001) state that a standard population is utilised for a number of years whereby the underlying age-structure of the population will change. Therefore, it is difficult to match a standard population to a population of interest as the underlying age-structure of the population of interest is forever changing. Ahmad et al. (2001) propose that selecting one standard population over another standard population because of matching age-structures is "insufficient justification" for the selection because of the changing age-structures. The standard population should be selected such that the average age-structure of the population(s) of interest are reflected over the time of the study in the standard population.

Data from Ahmad et al. (2001) and Sankoh et al. (2014) are combined to produce the following table:

Table 2.2: The standard populations for age groups showing different population structures.

Age Group	World Average	sub-Saharan Africa	Asia	Segi	Scandinavian
$0-4$	8.86	16.25	12.19	12	8
$5-9$	8.69	14.74	12.02	10	7
$10-14$	8.60	13.13	11.90	9	7
$15-19$	8.47	10.81	10.78	9	7
$20-24$	8.22	8.41	8.72	8	7
$25-29$	7.93	6.90	7.26	8	7
$30-34$	7.61	5.73	6.51	6	7
$35-39$	7.15	4.76	5.93	6	7
$40-44$	6.59	4.03	5.31	6	7
$45-49$	6.04	3.44	4.56	6	7
$50-54$	5.37	2.86	3.78	5	7
$55-59$	4.55	2.45	3.19	4	6
$60-64$	3.72	2.03	2.68	4	5
$65-69$	2.96	1.65	2.07	3	4
$70-74$	2.21	1.22	1.47	2	3
$75-79$	1.52	0.81	0.89	1	2
$80-84$	0.91	0.44	0.46	0.5	1
$85+$	0.63	0.35	0.28	0.5	1
Total	100	100	100	100	100

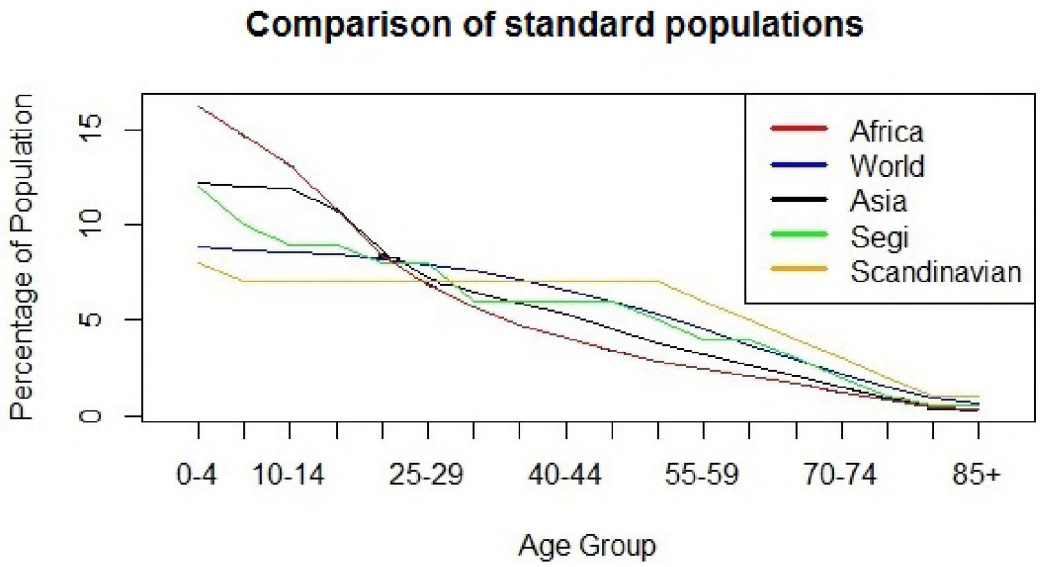


Figure 2.7: A comparison of standard populations.

When comparing the percentage of the population of the World and sub-Saharan Africa, it is easily seen that the lower age structures and higher age structures differ significantly. The World standard

population is thus inappropriate for many LMICs. Table 2.3 shows the differences between the World average and sub-Saharan Africa population structures.

Table 2.3: A comparison of the world average and sub-Saharan Africa population structures.

Age Group	World Average	sub-Saharan Africa	Difference
$0-4$	8.86	16.25	-7.39
$5-9$	8.69	14.74	-6.05
$10-14$	8.60	13.13	-4.53
$15-19$	8.47	10.81	-2.34
$20-24$	8.22	8.41	-0.19
$25-29$	7.93	6.90	1.03
$30-34$	7.61	5.73	1.88
$35-39$	7.15	4.76	2.39
$40-44$	6.59	4.03	2.56
$45-49$	6.04	3.44	2.60
$50-54$	5.37	2.86	2.51
$55-59$	4.55	2.45	2.10
$60-64$	3.72	2.03	1.69
$65-69$	2.96	1.65	1.31
$70-74$	2.21	1.22	0.99
$75-79$	1.52	0.81	0.71
$80-84$	0.91	0.44	0.47
$85+$	0.63	0.35	0.28
Total	**100**	**100**	

For this thesis, the standard populations are the mid-year population estimates for South Africa in 2014 and Georgia in 2004, as comparison is made over one population. The mid-year population estimates from Statistics South Africa (2014) for South Africa by age and sex in 2014 are shown in Table 2.4. The mid-year population estimates by county for Georgia are given in Tables B.1 and B.2 in Appendix B.1.

Table 2.4: The mid-year population estimates by age and sex for South Africa in 2014.

Age Group	Male	Female	Total
$0-4$	2892219	2827110	5719329
$5-9$	2692433	2644277	5336710
$10-14$	2580229	2543497	5123726
$15-19$	2624166	2593395	5217560
$20-24$	2662829	2604288	5267117
$25-29$	2515096	2439436	4951532
$30-34$	2034229	2056054	4090283
$35-39$	1739688	1763529	3503217
$40-44$	1482086	1639736	3121822
$45-49$	1270867	1482603	2753470
$50-54$	1089941	1287789	2377730
$55-59$	907807	1086583	1994390
$60-64$	703921	866257	1570178
$65-69$	492791	683675	1176466
$70-74$	327812	503451	831263
$75-79$	202623	328254	530877
$80+$	147270	286011	433281
Total	**26366008**	**27635944**	**54001953**

2.6 Conditional Independence

Lawson (2009) states that under a variety of conditions it may be possible to use the conditional independence of the data given the parameters at a "higher level of the hierarchy." Suppose the count data from the i^{th} region is denoted by y_i, then the count data may be assumed to be independent of other count data outcomes when given information about other model parameters. That is, let $\boldsymbol{\theta}$ be a parameter vector then the conditional $y_i|\boldsymbol{\theta}$ can be assumed to have an independent contribution to the experiment which means that "dependence only exists unconditionally" according to Lawson (2009). This method does not cover all situations because there can still be "residual correlation effects" after adding confounders. The method of hierarchical modelling is then introduced via conditional independence.

2.7 Hierarchical Modelling

A Bayesian hierarchical model is defined by Lesaffre & Lawson (2012) as a model with statistical procedures for data which has a hierarchical structure. This data is referred to as clustered data and is often correlated. A defining feature of hierarchical modelling is that the observed counts, denoted by y_{ij}, where the units are represented by i and within groups are represented by j are used to estimate

values of the population distribution of the θ_j when the values of θ_j are unknown as stated by Gelman et al. (2014). The hierarchical model prevents model overfitting by incorporating the population distribution into the model, this allows for some dependency between the parameters in the model. Non-hierarchical methods cannot incorporate many parameters into a model and there is little or no dependence between parameters. Thus the parameters do not always fit the data well which leads to model overfitting. The frequentist method to approach clustered data according to Lesaffre & Lawson (2012) is the generalised estimating equations (GEEs) method first used by Liang & Zeger (1986). The Bayesian method accounts for the randomness in the parameters by assigning a distribution to each parameter thus fixed and random effects do not have to be considered in the Bayesian methodology.

Each parameter in a hierarchical model has its own distribution and, according to Lawson (2009), the distributions are most often derived by the experimenter and control the values of each parameter. The posterior distribution is derived by the formula: $\pi(y|\theta) \propto L(\theta|y)\pi(\theta)$ where $\pi(\theta) \sim gamma(\alpha, \beta)$ is the prior distribution and likelihood given in Lawson (2009) as:

$$L(\theta|y) \propto \prod_{i=1}^{N} \left[\Big(e_i\theta\Big)^{yi} exp\Big(-e_i\theta\Big) \right].$$

The parameters of the prior distribution (α, β) may possess their own distributions, such parameters are regarded as stochastic. These parameters are called hyperparameters and they possess hyperprior distributions. The parameters may also be assigned a value by the experimenter; this is only when the experimenter has a good background in the field and has past experiences.

Lesaffre & Lawson (2012) outline two reasons for using the Bayesian hierarchical model over the equivalent frequentist method, these reasons are:

- The Bayesian hierarchical model considers all possible uncertainties in each model parameter and allows for prior knowledge to be incorporated into the model.
- The Markov Chain Monte Carlo (MCMC) method used to sample from the posterior distribution enables flexibility and reduces the effect of the parametric assumptions used in the frequentist methods.

Figure 2.8 represents an example of a hierarchical model.

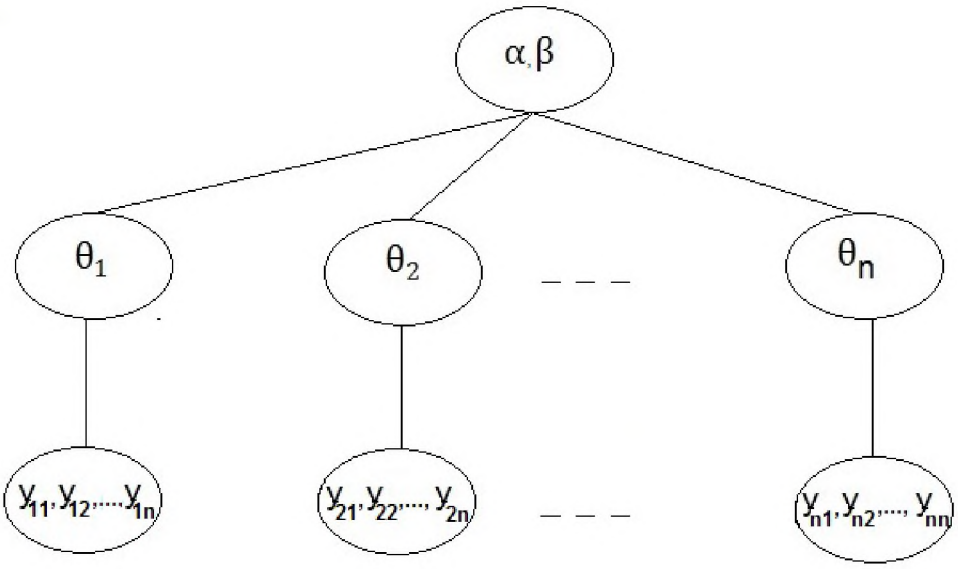


Figure 2.8: An example of a hierarchical model.

2.8 The Poisson Distribution

The Poisson distribution is a discrete probability distribution and is a special case of the binomial distribution. The Poisson distribution was formed by a French Mathematician called Siméon Denis Poisson (1781-1840) in 1837 in his work "*Recherches sur la probabilité des jugements en matiére criminelle et en matiére civile*" which translates to "Research on the Probability of Judgments in Criminal and Civil matters," according to West (2008). West (2008) states that Siméon Denis Poisson's work was formulated around random variables that involve positive count data in a specific time period, where the length of the interval is given.

Siméon Denis Poisson derived the Poisson distribution, which is a limiting factor of the binomial distribution, by using the Law of Small Numbers. Another contribution to the Poisson distribution was made by the Polish-German economist-statistician Ladislaus Bortkiewicz (1868-1931) who in 1898 investigated the number of deaths of soldiers in the Prussian army accidentally killed by horse kicks as stated by Selvi & Nishanthi (2012). King (1998) states that the Poisson probability distribution is derived using three principles.

The three principles are:

Firstly, suppose we have an observation denoted by i and we consider the time interval for the events in which observation i occurs. The total number of events which have occurred at the end of the time interval are denoted by Z_i which is thus also a random variable. The derivation of the Poisson distribution requires the formation of the events during the unobserved interval. The total number of events which have taken place in the time t during the period of observation i are denoted by the random variable Z_{ti}. The probability of an event occurring and not occurring is added to the total

number of events during the interval from t to $t + \triangle t$ and is given in King (1998) as:

$$P\Big(Z_{(t+\triangle t)i} = z_{ti} + 1 | Z_{ti} = z_{ti}\Big) = \lambda\triangle t + o(\triangle t) \text{ or} \tag{2.7}$$

$$P\Big(Z_{(t+\triangle t)i} = z_{ti} | Z_{ti} = z_{ti}\Big) = 1 - \{\lambda\triangle t + o(\triangle t)\}, \tag{2.8}$$

where $o(\triangle t)$ represents the probability that more than one event takes place in time $\triangle t$ and when it is divided by $\triangle t$, $\triangle t$ tends to 0 as $\triangle t$ becomes small, that is $\triangle t/\triangle t \to 0$ as $\triangle t$ gets small. The unconditional probability $P\Big(Z_{(t+\triangle t)i} = z_{ti} + 1\Big)$ can be written as two mutually exclusive events as stated by King (1998).

Secondly, the two mutually exclusive events are namely, the events z_{ti} which have taken place in the time period t and another event which takes place in the next $\triangle t$ interval and the events $z_{ti} + 1$ have all taken place during time t and no other events occur in the interval from t to $t + \triangle t$. The sum is a combination of Equations 2.7 and 2.8 and is then written by King (1998) as:

$$P\Big(Z_{(t+\triangle t)i} = z_{ti} + 1\Big) = P\Big(Z_{ti} = z_{ti}\Big)\lambda\triangle t + P\Big(Z_{ti} = z_{ti} + 1\Big)\Big(1 - \lambda\triangle t\Big). \tag{2.9}$$

For the derivation of Equation 2.9, two principles are required. The first one is that two events may not occur at exactly the same time hence $o(\triangle t)$ does not appear in Equation 2.9 and the second being that we assume the probability of an event taking place in the interval from t to $t + \triangle t$ is independent of all the other events taking place before time t. Thus Equation 2.9 can be represented as the product of the marginal probabilities as said by King (1998). The change in time for $P(Z_{ti} = z_{ti} + 1)$ is represented in Equations 2.10 and 2.11 for when $\triangle t$ gets small:

$$\frac{\partial P\Big(Z_{ti} = z_{ti} + 1\Big)}{\partial t} = \lim_{\triangle t \to 0} \frac{P\Big(Z_{(t+\triangle t)} = z_{ti} + 1\Big) - P\Big(Z_{ti} = z_{ti} + 1\Big)}{\triangle t} \tag{2.10}$$

$$= \begin{cases} \lambda\Big[P\Big(Z_{ti} = z_{ti}\Big) - P\Big(Z_{ti} = z_{ti} + 1\Big)\Big] & \text{when } z_{ti} + 1 > 1, \\ -\lambda P\Big(Z_{ti} = 0\Big) & \text{when } z_{ti} + 1 = 1. \end{cases} \tag{2.11}$$

The third and last principle is that no events have taken place before the start of the time interval, that is no events have taken place before $t = 0$, then King (1998) shows that $P\big(Z_{0i} = 0\big) = 1$ and then the distribution can be formed. Equation 2.11 is used and is solved as:

$$P\Big(Z_{ti} = 0\Big) = -\lambda^{-1}\frac{\partial P\Big(Z_{ti} = 0\Big)}{\partial t} = exp(-\lambda t), \tag{2.12}$$

which is a special case of the exponential distribution. King (1998) then substitutes Equation 2.12 into

Equation 2.11 which results in:

$$P\Big(Z_{ti} = 1\Big) = \lambda t e^{-\lambda t},$$

by solving this Equation and by substituting positive integers in place of Z_{ti}, the Poisson distribution is formed.

The Poisson probability density function (PDF), denoted by $f\Big(z_i|\lambda, t\Big)$, is well-known and given as:

$$f\Big(z_i|\lambda, t\Big) = \begin{cases} \frac{(\lambda t)^{z_i} e^{-\lambda t}}{z_i!} & \text{for when } t > 0,\ \lambda > 0 \text{ and } z_i = 0, 1, 2, \ldots \\ 0 & \text{Otherwise.} \end{cases}$$

This can also be written as:

$$f\Big(z_i|\lambda\Big) = \begin{cases} \frac{\lambda^{z_i} e^{-\lambda}}{z_i!} & \text{for when } \lambda > 0 \text{ and } z_i = 0, 1, 2, \ldots \\ 0 & \text{Otherwise.} \end{cases}$$

The cumulative distribution function (CDF) is well-known and given as:

$$F\Big(z_i|\lambda, t\Big) = e^{-\lambda} \sum_{i=0}^{\lfloor z_i \rfloor} \frac{\lambda^i}{i!} \text{ for } z_i = 0, 1, 2 \ldots$$

The expected value and variance of the Poisson distribution are well-known and given as:

$$\begin{aligned} E(Y) &= \lambda, \\ Var(Y) &= \lambda. \end{aligned}$$

Figure 2.9 represents the Poisson density function for different values of λ:

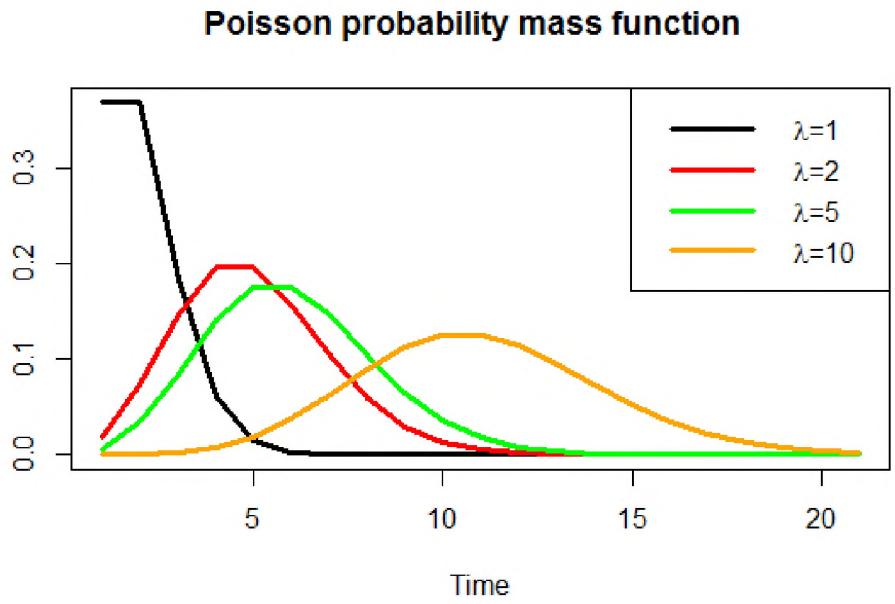


Figure 2.9: The Poisson density function.

Figure 2.10 shows the CDF for the Poisson distribution for different values of λ :

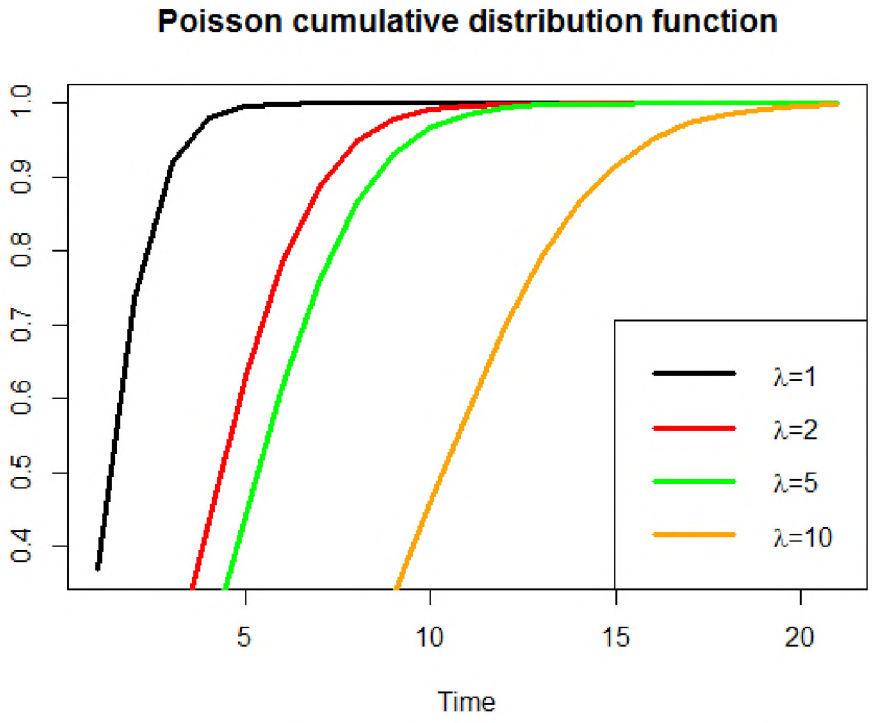


Figure 2.10: The Poisson cumulative distribution function.

An important property of the Poisson distribution according to Ross (2010) is that a Poisson random

variable may approximate a binomial random variable when the parameters of the binomial distribution are denoted by n and p respectively where n and p are small and for $n > 0$ and $0 \leq p \leq 1$. Ross (2010) shows this property using the following idea:

Consider a binomial random variable, denoted by Z, with parameters denoted by n and p respectively and let $\lambda = np$ then:

$$\begin{aligned} P\Big(Z = i\Big) &= \frac{n!}{\Big(n-i\Big)!i!}p^i\Big(1-p\Big)^{n-i} \\ &= \frac{n!}{\Big(n-i\Big)!i!}\left[\frac{\lambda}{n}\right]^i\left[1-\frac{\lambda}{n}\right]^{n-i} \\ &= \frac{n\Big(n-1\Big)\ldots\Big(n-i+1\Big)}{n^i}\frac{\lambda^i\Big(1-\frac{\lambda}{n}\Big)^n}{i!\Big(1-\frac{\lambda}{n}\Big)^i}. \end{aligned}$$

When n is large and p is small,

$$\begin{aligned} \left[1-\frac{\lambda}{n}\right]^n &\approx e^{-\lambda} \\ \frac{n\Big(n-1\Big)\ldots\Big(n-i+1\Big)}{n^i} &\approx 1 \\ \Big(1-\frac{\lambda}{n}\Big)^i &\approx 1. \end{aligned}$$

Therefore,

$$P\Big(Z = i\Big) \approx e^{-\lambda}\frac{\lambda^i}{i!}.$$

When dealing with count data suppose that $y_1, y_2, \ldots, y_N$ are the observed events of interest and assumed to be independent Poisson random variables with parameter $\mu_i,\ i = 1, 2, \ldots, N$. Suppose the model has $x_i,\ i = 1, 2, \ldots, N$ explanatory variables, then the model is written by Osei (2010) as:

$$\mu_i = x_i'\beta.$$

This model, however, may result in any real number but the value of μ_i must be non-negative. A logarithm is thus introduced into the model to allow μ_i to only assume non-negative values. Osei (2010) names this as the canonical link, denoted by $\eta_i = log\big(\mu_i\big)$ such that $\eta_i = x_i'\beta$. The resulting generalised linear model (GLM) with canonical link is:

$$log\Big(\mu_i\Big) = x_i'\beta,$$

where β is the "expected change in the log of the mean per unit change in the predictor variable, x_i." The likelihood for this model is:

$$L\Big(y_i|x_i\Big) = \prod_{i=1}^{N} \frac{exp\Big(-\mu_i\Big)\mu_i^{y_i}}{y_i!},$$

where $\mu_i = exp\Big(x_i'\beta\Big)$. The log likelihood is then:

$$l\Big(y_i|x_i\Big) = \sum_{i=1}^{N} \Big\{-\mu_i + y_i log\Big(\mu_i\Big) - log\Big(y_i!\Big)\Big\}.$$

Note: The above equations are just general for a Poisson distribution with parameter λ, slightly different notation will be used in this thesis, but will be defined accordingly.

2.8.1 Problems with the Poisson Distribution when using Count Data

2.8.1.1 Overview

The Poisson regression model is used by Berk & MacDonald (2008) when an experiment involves count data. The Poisson regression model is often used, as the formulation and application of the model is relatively simple. Techniques and methods which are used in a normal regression model are also valid for the Poisson regression model which makes it more appealing than other competing models. An important characteristic of the Poisson distribution is that the mean and variance, λ, are both equal. In count data, the mean and variance of the Poisson regression model should therefore be the same, but in practice this is not always the case.

2.8.1.2 Overdispersion

Overdispersion is defined by Yang et al. (2007) to occur when data has a higher variance than the mean and the characteristic of equal mean and variance in the Poisson distribution is therefore violated. This is often the case in count data. Overdispersion in a model means that the standard errors are underestimated and incorrect inference is made on each regression parameter. There are numerous methods to deal with overdispersion, one of the most common is named in Yang et al. (2007) as the generalised Poisson (GP) distribution.

Tests are available to determine whether a model shows overdispersion which includes, the likelihood ratio (LR) test, the Wald test and the score test.

2.8.1.3 Underdispersion

Underdispersion is thought by Giuffré et al. (2011) to be less common in count data than overdispersion. Underdispersion occurs when the mean of the Poisson regression model is higher than the variance of the model. This leads to bias in the model. Many factors may cause underdispersion but one of the more common factors is a repulsion in the data as proposed by Kokonendji et al. (2008).

Underdispersion may be detected by the Fisher index, which is the proportion of the variance and the mean. A Fisher index of 1 indicates that the data has an equal mean-variance relationship and an index of less than or greater than 1 indicates that there is underdispersion or overdispersion in the data.

2.9 Bayesian Statistics

Bayesian statistics is based on the concept of estimating an unknown state of nature with a probability. Bolstad (2004) states that the Bayes' theorem is used for estimation in Bayesian statistics and it allows one to assume a parameter is a random variable. The basic concepts of Bayesian statistics were first described by Thomas Bayes (1710-1761) who, in his essay, described the Bayes' theorem of inverse probability, along with other concepts. He described how the understanding of an experimenter (subjective belief) could be used in statistical inference. This description lead to what is known today as prior and posterior belief. After an experiment has run and there are results, the prior belief is updated and called the posterior belief. The essay, by Thomas Bayes, was only published in 1763, two years after the death of Thomas, nevertheless Bayesian Statistical Sciences have been further developed based on applications of his theorem as explained by Press (2009). Later on, other approaches to Bayesian statistics were developed.

The well-known Bayes' theorem is given by:

Let A and $B_1, \ldots, B_n$ be events where each B_i is disjoint, $\bigcup_{i=1}^{n} B_i = \Omega$, where Ω is the sample space, and $P\Big(B_i\Big) > 0$ for all i. Then:

$$P\Big(B_j|A\Big) = \frac{P\Big(A|B_j\Big)P\Big(B_j\Big)}{\sum_{i=1}^{n} P\Big(A|B_j\Big)P\Big(B_i\Big)}. \tag{2.13}$$

When an experiment has values of $\boldsymbol{H} = \Big(H_1, H_2, \ldots, H_n\Big)$ where $\boldsymbol{H}$ are unknown and an experimenter has a prior belief which can be expressed in terms of the prior frequency function $P(\boldsymbol{H})$ and there exists some random variables denoted by $\boldsymbol{E} = \Big(E_{n1}, E_{n2}, \ldots, E_{nk}\Big)$ which form a random sample from $P(\boldsymbol{E}|\boldsymbol{H})$ as in Lee (2004). The joint distribution in this experiment can be in the form of:

$$P(\boldsymbol{E}|\boldsymbol{H}) = P\Big(E_{n1}, E_{n2}, \ldots, E_{nk}|\boldsymbol{H}\Big) = P\Big(E_{n1}|\boldsymbol{H}\Big)P\Big(E_{n2}|\boldsymbol{H}\Big)\ldots P\Big(E_{nk}|\boldsymbol{H}\Big). \tag{2.14}$$

Then using Equations 2.13 and 2.14 to obtain the posterior distribution as in Lee (2004):

$$P(\boldsymbol{H}|\boldsymbol{E}) = \frac{P(\boldsymbol{E}|\boldsymbol{H})P(\boldsymbol{H})}{\int P(\boldsymbol{E}|\boldsymbol{H})P(\boldsymbol{H})dH} = \frac{P(\boldsymbol{E}|\boldsymbol{H})P(\boldsymbol{H})}{P(\boldsymbol{E})}.$$

Hence, the following relation is made:

$$P(\boldsymbol{H}|\boldsymbol{E}) = cP(\boldsymbol{E}|\boldsymbol{H})P(\boldsymbol{H}) \text{ where } c \text{ is a constant and } L(\boldsymbol{H}|\boldsymbol{E}) = P(\boldsymbol{E}|\boldsymbol{H}).$$

And then,

$$P(\boldsymbol{H}|\boldsymbol{E}) \propto P(\boldsymbol{E}|\boldsymbol{H})P(\boldsymbol{H}).$$

These equations give the relation:

$$posterior \propto prior \times likelihood.$$

The Bayes' theorem does not give one an indication of what prior to use when conducting an experiment.

2.9.1 Priors

2.9.1.1 Objective Prior

According to Press (2009), Thomas Bayes used the Bayes' theorem when some data had a binomial distribution with an unknown parameter p; this implies that he used a prior in which all outcomes had an equal probability. This prior is an objective prior.

Objectivity, more commonly known as frequency probability, is defined in Press (2009) as when an experimenter has very little or no knowledge prior to an experiment. There is also very little known about the unknown parameters in the experiment. Objective priors are limited to experiments which are repeated a large number of times. The advantages of using objective priors are defined in Press (2009) as:

- Many experiments have little information known prior to the experiment.
- Experiments repeated by different researchers result in similar results, unlike the subjective prior experiments.
- Objective priors are often used in public policy priors.

The disadvantages of using objective priors are stated in Press (2009) as:

- The parameters are assumed to be independent when in fact they are dependent and have a significant correlation.

- The objective prior can produce results which are undesirable in inference.
- A reader cannot see the experimenters' state of mind at the time of defining the prior.
- Improper priors sometimes cause indeterminacies when determining Bayes' factors.

2.9.1.2 Subjective Prior

The term subjective is often used in Bayesian statistics and is defined by Press (2009) as a personal opinion of an individual. Press (2009) states that the Bayes' theorem gives the extent of subjectivity about some unknown event after it has occurred and then gives the relation of two types of information.

The one type of information separates the subjects that have been observed this is called the "objective portion" of posterior belief. The other type of information is the extent of an individual's subjectivity about an unknown event. This may be skewed by past experiences or past experiments with these events. Subjectivity is important in Bayesian statistics since an experiments' results may be influenced by an experimenters personal opinions. The advantages of using subjective priors, according to Press (2009) include:

- The subjective prior distribution always integrates to one in the continuous case and sums to one in the discrete case.
- The subjective prior does not have a big effect on the posterior distribution.
- The subjective prior gives the reader an understanding of what the experimenter was thinking at the time of the experiment as well as a look into the state of mind of the experimenter.
- There is often insufficient information available to use the objective prior.

The disadvantages of using subjective priors are included in Press (2009) as :

- It may be difficult in some experiments to assess the subjective prior distribution which results in a meaningless probability distribution. Hyperparameters are created when a parameter indexes a subjective prior, these may also be difficult to assess.
- Since each experimenter may have a different subjective prior, the results of one experiment may not be meaningful to another experimenter who uses a different subjective prior.
- Sometimes the subjective prior is difficult to solve mathematically.

2.9.1.3 Reference Prior

The idea of the reference prior was first described in a paper written by Bernardo (1979) and adapted by Jim Berger and other authors as stated in Jordan (2010). The reference prior is a prior distribution with the aim of maximising a chosen measure of distance between the posterior distribution and the prior distribution as data are observed. There are many measures of distances which can be chosen by the experimenter and may be dependent on the type of experiment and the data used in the experiment. This is done because an experimenter wants the data of the experiment to have a maximum effect on the posterior distribution. The Jeffreys' prior is equivalent to the reference prior in the one-dimensional parameter case as explained by Jordan (2010).

The reference prior, denoted by $P(\alpha)$, is determined by taking the expectation of the measure of distance given the distribution of a model for the data in the experiment. This is expressed in Jordan (2010) as:

$$P^{\star}(\alpha) = \underset{P(t)}{argmax} I(\alpha, T),$$

where

$$I(\alpha, T) = \int P(t) \int P(\alpha|t) log\left(\frac{P(\alpha|t)}{P(\alpha)}\right) d\alpha dt = \int \int P(\alpha, t) log\left(\frac{P(\alpha, t)}{P(\alpha)P(t)}\right) d\alpha dt.$$

2.9.1.4 Other Priors

The method used by Robert (2007) is to check for good frequentist properties but it is very unusual to derive such a prior. Frequentist properties are those which are maintained on average, rather than being conditional on a variable. Sun (1997) states that matching priors are important noninformative priors and these matching priors lead to posterior confidence regions. The noninformative prior is used when little or no information is available to an experimenter prior to an experiment. In some cases the likelihood rules over the prior significantly, this may be the case for two reasons. The first reason is that researchers' may have two different prior beliefs about an experiment, which may result in different outcomes for the same experiment. Then it is reasonable to use a reference prior which is an account of both researchers' prior beliefs and is ruled by the likelihood. The other reason is that the results of any experiment are meant to increase the knowledge of a reader. Otherwise the experiment is a failure, if this is the case then the likelihood will dominate the prior as explained by Lee (2004). This noninformative prior for ϕ is denoted by: $\pi(\phi) = c$,where $0 < \phi < \infty$ and c is some constant. This is not a proper density since it does not integrate to one.

In the next section, various priors will be given for $y_i \sim Pois\Big(\theta_i\Big)$. The priors will be derived for a general case of $y_1, y_2, \ldots, y_N$ independent Poisson distributions.

2.9.1.5 Probability Matching Prior

The probability matching prior (PMP) is defined in Datta & Sweeting (2005) as a prior distribution where the posterior probabilities of some region are similar to the coverage probabilities, either by the exact value of the probability or by the approximate value of the probability. In very rare or limited cases the probability matching prior has an exact coverage probability. The probability matching prior is widely considered as a nonsubjective prior. This prior is widely used as the data does not have to be identically and independently distributed (IID) as the prior can be extended to cover the dependent case. A matching prior is second-order when the coverage probability is not similar to the credible level by values which are of the order n^{-1}, as explained in Datta & Sweeting (2005). The algorithm below is valid for location families and one-sided confidence intervals and is given by Raubenheimer & Van der Merwe (2014) as:

- Derive the likelihood function for the vector $\boldsymbol{\theta} = \Big[\theta_1, \theta_2, \ldots, \theta_k\Big]$ and denote the likelihood function by $L\Big(\boldsymbol{\theta}|y_1, y_2, \ldots, y_N\Big)$ where $\boldsymbol{\theta}$ is unknown.
- Derive the inverse Fisher information matrix, $I^{-1}(\boldsymbol{\theta})$, for $\boldsymbol{\theta}$.
- Suppose the probability matching prior is denoted by $t(\boldsymbol{\theta})$, then derive:
  1. $\nabla_t'(\boldsymbol{\theta}) = \Big[\frac{\partial t(\boldsymbol{\theta})}{\partial \theta_1} \; \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_2} \; \cdots \; \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_k}\Big]$
  2. $\nabla_t(\boldsymbol{\theta}) = \Big[\frac{\partial t(\boldsymbol{\theta})}{\partial \theta_1} \; \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_2} \; \cdots \; \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_k}\Big]'$.
- Let $\eta'(\boldsymbol{\theta}) = \frac{\nabla_t'(\boldsymbol{\theta}) I^{-1}(\boldsymbol{\theta})}{\sqrt{\nabla_t'(\boldsymbol{\theta}) I^{-1}(\boldsymbol{\theta}) \nabla_t(\boldsymbol{\theta})}}$.
- When the differential equation $\sum_{j=1}^{k} \frac{\partial}{\partial \theta_i} \{\eta_i(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\} = 0$ holds then $\pi(\boldsymbol{\theta})$ is a probability matching prior.

Theorem 2.1. *The probability matching prior for the Poisson distribution is:*

$$\pi_{PM}(\boldsymbol{\theta}) \propto \left[\sum_{i=1}^{n} b_i^2 \theta_i^{-1}\right]^{\frac{1}{2}}.$$

Proof. Let $t(\boldsymbol{\theta})$ denote the product of n Poisson rates with ranging powers, then $t(\boldsymbol{\theta}) = \prod_{i=1}^{n} \theta_i^{b_i}$.

The Poisson PDF will then have the form:

$$f\Big(y_i|\theta_i\Big) = \frac{\theta_i^{y_i} exp\Big(-\theta_i\Big)}{y_i!} \text{ for } y_i = 0, 1, 2, \ldots$$

The likelihood for this distribution is:

$$L\Big(\theta_i|y_i\Big) \propto \prod_{i=1}^{n} \frac{\theta_i^{y_i} exp\Big(-\theta_i\Big)}{y_i!},$$

which results in a log likelihood as:

$$l\Big(\theta_i|y_i\Big) = Constant + \sum_{i=1}^{n} y_i log\Big(\theta_i\Big) - \sum_{i=1}^{n} \theta_i.$$

Differentiation of the log likelihood with respect to θ_i yields:

$$\frac{\partial l\Big(\theta_i|y_i\Big)}{\partial \theta_i} = \frac{y_i}{\theta_i} - 1 \text{ where } i = 1, 2, \ldots, n$$

and the second derivatives are:

$$\begin{aligned} \frac{\partial^2 l\Big(\theta_i|y_i\Big)}{\partial \theta_i^2} &= -\frac{y_i}{\theta_i^2} \\ \frac{\partial^2 l\Big(\theta_i|y_i\Big)}{\partial \theta_i \partial \theta_j} &= 0 \text{ where } i \neq j. \end{aligned}$$

With expectation of the second derivatives:

$$\begin{aligned} -E\left[\frac{\partial^2 l\Big(\theta_i|y_i\Big)}{\partial \theta_i^2}\right] &= -E\left[-\frac{y_i}{\theta_i^2}\right] \\ &= \frac{1}{\theta_i^2} E\Big[y_i\Big] \\ &= \frac{1}{\theta_i^2}\theta_i \\ &= \theta_i^{-1}. \end{aligned}$$

The Fisher matrix and inverse Fisher matrix are:

$$
\begin{aligned}
F(\boldsymbol{\theta}) &= \begin{bmatrix} \frac{1}{\theta_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\theta_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\theta_n} \end{bmatrix} \\
F^{-1}(\boldsymbol{\theta}) &= \begin{bmatrix} \theta_1 & 0 & 0 & \dots & 0 \\ 0 & \theta_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \theta_n \end{bmatrix}.
\end{aligned}
$$

Let

$$
\begin{aligned}
\nabla' t(\boldsymbol{\theta}) &= \left[\frac{\partial t(\boldsymbol{\theta})}{\partial \theta_1} \; \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_2} \dots \frac{\partial t(\boldsymbol{\theta})}{\partial \theta_n} \right] \\
&= \left[b_1 \theta_1^{b_1 - 1} \prod_{i \neq 1}^{n} \theta_i^{b_i} \; b_2 \theta_2^{b_2 - 1} \prod_{i \neq 2}^{n} \theta_i^{b_i} \dots b_n \theta_n^{b_n - 1} \prod_{i \neq n}^{n} \theta_i^{b_i} \right],
\end{aligned}
$$

then,

$$
\begin{aligned}
\nabla' t(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) &= \left[b_1 \theta_1^{b_1 - 1} \prod_{i \neq 1}^{n} \theta_i^{b_i} \; b_2 \theta_2^{b_2 - 1} \prod_{i \neq 2}^{n} \theta_i^{b_i} \dots b_n \theta_n^{b_n - 1} \prod_{i \neq n}^{n} \theta_i^{b_i} \right] \begin{bmatrix} \theta_1 & 0 & 0 & \dots & 0 \\ 0 & \theta_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \theta_n \end{bmatrix} \\
&= \left[b_1 \theta_1^{b_1} \prod_{i \neq 1}^{n} \theta_i^{b_i} \; b_2 \theta_2^{b_2} \prod_{i \neq 2}^{n} \theta_i^{b_i} \dots b_n \theta_n^{b_n} \prod_{i \neq n}^{n} \theta_i^{b_i} \right]
\end{aligned}
$$

$$\begin{aligned}
\nabla' t(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) \nabla t(\boldsymbol{\theta}) &= \left[b_1 \theta_1^{b_1} \prod_{i \neq 1}^{n} \theta_i^{b_i} \; b_2 \theta_2^{b_2} \prod_{i \neq 2}^{n} \theta_i^{b_i} \ldots b_n \theta_n^{b_n} \prod_{i \neq n}^{n} \theta_i^{b_i} \right] \begin{bmatrix} b_1 \theta_1^{b_1 - 1} \prod_{i \neq 1}^{n} \theta_i^{b_i} \\ b_2 \theta_2^{b_2 - 1} \prod_{i \neq 2}^{n} \theta_i^{b_i} \\ \vdots \\ b_n \theta_n^{b_n - 1} \prod_{i \neq n}^{n} \theta_i^{b_i} \end{bmatrix} \\
&= b_1^2 \theta_1^{b_1 - 1} \prod_{i \neq 1}^{n} \theta_i^{b_i} \prod_{i=1}^{n} \theta_i^{b_i} + b_2^2 \theta_2^{b_2 - 1} \prod_{i \neq 2}^{n} \theta_i^{b_i} \prod_{i=1}^{n} \theta_i^{b_i} + \ldots + b_n^2 \theta_n^{b_n - 1} \prod_{i \neq n}^{n} \theta_i^{b_i} \prod_{i=1}^{n} \theta_i^{b_i} \\
&= \left(\sum_{i=1}^{n} b_i^2 \right) \left(\prod_{i=1}^{n} \theta_i^{b_i} \right) \left(\theta_1^{b_1 - 1} \prod_{i \neq 1}^{n} \theta_i^{b_i} + \theta_2^{b_2 - 1} \prod_{i \neq 2}^{n} \theta_i^{b_i} + \ldots + \theta_n^{b_n - 1} \prod_{i \neq n}^{n} \theta_i^{b_i} \right) \\
&= \left(\sum_{i=1}^{n} \frac{b_i^2}{\theta_i} \right) \left(\prod_{i=1}^{n} \theta_i^{b_i} \right)^2 .
\end{aligned}$$

Let

$$\begin{aligned}
\eta'(\boldsymbol{\theta}) &= \frac{\nabla' t(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta})}{\sqrt{\nabla' t(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) \nabla t(\boldsymbol{\theta})}} \\
&= \frac{\left[b_1 \theta_1^{b_1} \prod_{i \neq 1}^{n} \theta_i^{b_i} \; b_2 \theta_2^{b_2} \prod_{i \neq 2}^{n} \theta_i^{b_i} \ldots b_n \theta_n^{b_n} \prod_{i \neq n}^{n} \theta_i^{b_i} \right]}{\sqrt{\left(\sum_{i=1}^{n} \frac{b_i^2}{\theta_i} \right) \left(\prod_{i=1}^{n} \theta_i^{b_i} \right)^2}} \\
&= \left[\frac{b_1}{\sqrt{\left(\sum_{i=1}^{n} \frac{b_i^2}{\theta_i} \right)}} \quad \frac{b_2}{\sqrt{\left(\sum_{i=1}^{n} \frac{b_i^2}{\theta_i} \right)}} \quad \cdots \quad \frac{b_n}{\sqrt{\left(\sum_{i=1}^{n} \frac{b_i^2}{\theta_i} \right)}} \right] \\
&= \left[\eta_1(\boldsymbol{\theta}) \quad \eta_2(\boldsymbol{\theta}) \quad \ldots \quad \eta_n(\boldsymbol{\theta}) \right].
\end{aligned}$$

Let the prior, $\pi(\boldsymbol{\theta})$, be given as:

$$\pi(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\frac{b_i^2}{\theta_i} \right)^{\frac{1}{2}},$$

then $\pi(\boldsymbol{\theta})$ is a PMP if and only if $\sum_{i=1}^{n} \frac{\partial}{\partial \theta_i}\left[\eta_i(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\right] = 0$.

$$\begin{aligned}
\eta_i(\boldsymbol{\theta})\pi(\boldsymbol{\theta}) &= \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} \\
\frac{\partial}{\partial \theta_1}\left[\eta_1(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\right] &= \frac{\partial}{\partial \theta_1} b_1 = 0 \\
\frac{\partial}{\partial \theta_2}\left[\eta_2(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\right] &= \frac{\partial}{\partial \theta_2} b_2 = 0 \\
&\vdots \\
\frac{\partial}{\partial \theta_n}\left[\eta_n(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\right] &= \frac{\partial}{\partial \theta_n} b_n = 0 \\
\therefore \sum_{i=1}^{n} \frac{\partial}{\partial \theta_i}\left[\eta_i(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\right] &= 0.
\end{aligned}$$

Thus $\pi_{PM}(\boldsymbol{\theta}) \propto \left[\sum_{i=1}^{n} b_i^2 \theta_i^{-1}\right]^{\frac{1}{2}}$ is a probability matching prior. □

2.9.1.6 Jeffreys' Prior

Box & Tiao (1992) give the following motivation for using the Jeffreys' prior:

- The Jeffreys' prior is invariant under monotone transformations.
- The Jeffreys' prior can be extended to the multiparameter case.
- There is a lack of information in an experiment.
- The Jeffreys' prior is locally uniform and noninformative.

Kass & Wasserman (1996) propose a set of formal rules to select a prior distribution based on the rules previously outlined by Jeffreys'. Rules for using a prior distribution in a problem which involves estimation involves assigning an equal probability to each parameter value. When the parameter space has a bounded interval, then the prior density is determined to be constant. Another case involves using an improper prior. The last case is used when the standard deviation is unknown and this prior is invariant to power transformations of the parameters as stated by Kass & Wasserman (1996).

Theorem 2.2. *The Jeffreys' prior for the Poisson distribution is:*

$$\pi_J(\boldsymbol{\theta}) \propto \left(\prod_{i=1}^{n} \theta_i\right)^{-\frac{1}{2}}.$$

Proof. The Fisher information matrix for the Poisson distribution is given by:

$$F(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\theta_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\theta_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\theta_n} \end{bmatrix}.$$

Then the Jeffreys' prior is $\sqrt{F(\boldsymbol{\theta})}$ which is:

$$\sqrt{F(\boldsymbol{\theta})} = \begin{bmatrix} \theta_1^{-\frac{1}{2}} & 0 & 0 & \dots & 0 \\ 0 & \theta_2^{-\frac{1}{2}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \theta_n^{-\frac{1}{2}} \end{bmatrix}.$$

Thus the Jeffreys' prior is:

$$\pi_J(\boldsymbol{\theta}) \propto \left(\prod_{i=1}^{n} \theta_i \right)^{-\frac{1}{2}}.$$

□

2.9.1.7 Conjugate Prior

A conjugate prior is a prior distribution which comes from the same family as the posterior distribution. Conjugate priors are often used because of simplicity since the posterior distribution will always have a closed form and inference will correspond to that of the prior distribution. Lawson (2009) shows that a conjugate prior is derived by observing the kernel of the relationship between the "prior-likelihood product."

A conjugate prior may not work in a hierarchical model with a high number of parameters, but Lawson (2009) uses conditional conjugacy to check if a model is adequate.

Theorem 2.3. *The conjugate prior associated with a single Poisson likelihood is:*

$$\pi_C(\theta) \propto \theta^{\alpha-1} exp(-\beta\theta) \sim gamma(\alpha, \beta).$$

Proof. Since the posterior distribution is:

$$\pi(\theta|y) \propto L(\theta|y)\pi(\theta),$$

where

$$\pi(\theta) \propto \theta^{\alpha-1} exp(-\beta\theta),$$

and

$$L(\theta|y) \propto \prod_{i=1}^{n} \theta^{y_i} exp(-\theta).$$

Then:

$$\begin{aligned}
\pi_C(\theta|y) &\propto \theta^{\alpha-1} exp(-\beta\theta) \prod_{i=1}^{n} \theta^{y_i} exp(-\theta) \\
&\propto \theta^{\alpha+\sum_{i=1}^{n} yi-1} exp(-\beta\theta - n\theta) \\
&\propto \theta^{\alpha+\sum_{i=1}^{n} yi-1} exp\Big(-\theta(\beta+n)\Big) \\
&\sim gamma\left(\alpha+ \sum_{i=1}^{n} yi, \beta+n\right).
\end{aligned}$$

Since the posterior distribution comes from the same family as the prior distribution, the prior distribution $\pi_C(\theta) \propto \theta^{\alpha-1} exp(-\beta\theta)$ is a conjugate prior. □

2.9.1.8 Divergence Prior

The divergence prior is equal to the absolute value of the fourth root of the Fisher information number as in Ghosh et al. (2011). This prior is a "unique optimising prior" and makes use of the Kullback-Leibler (KL) distance between the prior distribution and the associated posterior distribution and the Bhattacharyya-Hellinger divergence proposed by Bhattacharyya (1943) and Hellinger (1909). The proof of the divergence prior can be found in Ghosh et al. (2011).

Theorem 2.4. *The divergence prior for the Poisson distribution is:*

$$\pi_D(\boldsymbol{\theta}) \propto \left(\prod_{i=1}^{n} \theta_i\right)^{-\frac{1}{4}}.$$

Proof. The Fisher information matrix for the Poisson distribution is given by:

$$F(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\theta_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\theta_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\theta_n} \end{bmatrix}.$$

The absolute value of the fourth root of the Fisher information matrix is then:

$$\sqrt[4]{F(\boldsymbol{\theta})} = \begin{bmatrix} \theta_1^{-\frac{1}{4}} & 0 & 0 & \dots & 0 \\ 0 & \theta_2^{-\frac{1}{4}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \theta_n^{-\frac{1}{4}} \end{bmatrix}.$$

Thus the divergence prior is:

$$\pi_D(\boldsymbol{\theta}) \propto \left(\prod_{i=1}^{n} \theta_i \right)^{-\frac{1}{4}}.$$

□

2.9.2 Empirical Bayes

Empirical Bayes is said by Marshall (1991) to reduce the mean square error by pooling estimates over the regions in the study. Let the area under study be made up of N regions. Suppose that the events in the study are measured over several years. Let $\ominus_i$ be the event rate per annum at the i^{th} region. Suppose that the cumulative events in the i^{th} region, denoted by r_i, are Poisson random variables with conditional mean given by Marshall (1991) as:

$$E\Big(r_i|\ominus_i\Big) = k_i \ominus_i,$$

where k_i is the number of "person-years at risk." The ratio of $x_i = \frac{r_i}{k_i}$ is a "crude estimator" of $\ominus_i$ and is also the maximum likelihood of $\ominus_i$. x_i has a mean $E\big(x_i|\ominus_i\big) = \ominus_i$ which is conditional on $\ominus_i$ and variance $var\big(x_i|\ominus_i\big) = \frac{\ominus_i}{k_i}$ also conditional.

Suppose that r_i are assumed to be independent and $\ominus_i$ has a prior distribution with parameters $C_i = E_{\ominus}\big(\ominus_i\big)$ and $L_i = var_{\ominus}\big(\ominus_i\big)$ for the mean and variance respectively. The unconditional mean and variance of x_i are given in Marshall (1991) as:

$$E_x\Big(x_i\Big) = E_{\ominus}\Big(E\big(x_i|\ominus_i\big)\Big) = E_{\ominus}(\ominus_i) = C_i,$$

and unconditional variance

$$\begin{aligned} Var_x\Big(x_i\Big) &= var_{\ominus}\Big(E\big(x_i|\ominus_i\big)\Big) + E_{\ominus}\Big(var\big(x_i|\ominus_i\big)\Big) \\ &= var_{\ominus}\Big(\ominus_i\Big) + E_{\ominus}\left(\frac{\ominus_i}{k_i}\right) \\ &= L_i + \frac{C_i}{k_i}. \end{aligned}$$

Applying the shrinkage estimator derived by Efron & Morris (1973), the best linear Bayes estimator of $\ominus_i$ when the values of C_i and L_i are known is:

$$\hat{\ominus}_i = C_i + F_i\Big(x_i - C_i\Big),$$

where $F_i = \dfrac{L_i}{\left(L_i + \frac{C_i}{k_i}\right)} = \dfrac{var_{\ominus}\Big(\ominus_i\Big)}{var_x\Big(x_i\Big)}$, which is the proportion of the prior variance of $\ominus_i$ to the unconditional variance of x_i.

2.9.3 Posterior Distribution

The posterior distribution is derived by:

$$\pi(\theta|y) \propto L(\theta|y)\pi(\theta).$$

The posterior distributions in this section will be given for a single Poisson likelihood.

Theorem 2.5. *The posterior distribution using the Jeffreys' prior and Poisson likelihood is:*

$$\pi_J(\theta|y) \propto \theta^{\sum_{i=1}^{n} y_i - \frac{1}{2}} exp(-n\theta) \sim gamma\left(\sum_{i=1}^{n} y_i + \frac{1}{2}, n\right).$$

Proof. Let $\pi_J(\theta) \propto \theta^{-\frac{1}{2}}$ then the posterior distribution is:

$$\begin{aligned}
\pi_J(\theta|y) &\propto \prod_{i=1}^{n} \frac{\theta^{y_i} exp(-\theta)}{y_i!}\left(\theta^{-\frac{1}{2}}\right) \\
&\propto \theta^{\sum_{i=1}^{n} y_i - \frac{1}{2}} exp(-n\theta) \\
&\sim gamma\left(\sum_{i=1}^{n} y_i + \frac{1}{2}, n\right).
\end{aligned}$$

□

Theorem 2.6. *The posterior distribution using the probability matching prior and Poisson likelihood is:*

$$\pi_{PM}(\theta|y) \propto \theta^{\sum_{i=1}^{n} y_i - \frac{n}{2}} exp(-n\theta) \sim gamma\left(\sum_{i=1}^{n} y_i + \frac{1}{2}, n\right).$$

Proof. Let $\pi_{PM}(\theta) \propto \left(\theta^{-1}\right)^{\frac{1}{2}}$, by letting $b_i = 1$ and considering a single likelihood, then the posterior

distribution is:

$$\begin{aligned}
\pi_{PM}(\theta|y) &\propto \prod_{i=1}^{n} \frac{\theta^{y_i} exp(-\theta)}{y_i!} \left(\theta^{-\frac{1}{2}}\right) \\
&\propto \prod_{i=1}^{n} \theta^{y_i} exp(-\theta) \left(\theta^{-\frac{1}{2}}\right) \\
&\propto \theta^{\sum_{i=1}^{n} y_i - \frac{1}{2}} exp(-n\theta) \\
&\sim gamma\left(\sum_{i=1}^{n} y_i + \frac{1}{2}, n\right).
\end{aligned}$$

□

Theorem 2.7. *The posterior distribution using the divergence prior and Poisson likelihood is:*

$$\pi_D(\theta|y) \propto \theta^{\sum_{i=1}^{n} y_i - \frac{1}{4}} exp(-n\theta) \sim gamma\left(\sum_{i=1}^{n} y_i + \frac{3}{4}, n\right).$$

Proof. Let $\pi_D(\theta) \propto \theta^{-\frac{1}{4}}$ then the posterior distribution is:

$$\begin{aligned}
\pi_D(\theta|y) &\propto \prod_{i=1}^{n} \frac{\theta^{y_i} exp(-\theta)}{y_i!} \left(\theta^{-\frac{1}{4}}\right) \\
&\propto \theta^{\sum_{i=1}^{n} y_i - \frac{1}{4}} exp(-n\theta) \\
&\sim gamma\left(\sum_{i=1}^{n} y_i + \frac{3}{4}, n\right).
\end{aligned}$$

□

Theorem 2.8. *The posterior distribution using a conjugate prior and Poisson likelihood is:*

$$\pi_C(\theta|y) \propto \theta^{\alpha + \sum_{i=1}^{n} yi - 1} exp\big(-\theta(\beta + n)\big) \sim gamma\left(\alpha + \sum_{i=1}^{n} yi, \beta + n\right).$$

Proof. Let $\pi_C(\theta) \propto \theta^{\alpha-1}exp(-\beta\theta) \sim gamma(\alpha, \beta)$ then the posterior distribution is:

$$\begin{aligned}
\pi_C(\theta|y) &\propto \theta^{\alpha-1}exp(-\beta\theta)\prod_{i=1}^{n}\theta^{y_i}exp(-\theta) \\
&\propto \theta^{\alpha+\sum_{i=1}^{n}yi-1}exp(-\beta\theta - n\theta) \\
&\propto \theta^{\alpha+\sum_{i=1}^{n}yi-1}exp\big(-\theta(\beta+n)\big) \\
&\sim gamma\left(\alpha+\sum_{i=1}^{n}yi, \beta+n\right).
\end{aligned}$$

□

Properness of a density requires two properties to be satisfied. The first property is the density must integrate to one in the continuous case and sum to one in the discrete case. The second property is $f(t) \geq 0$, for all t. For all these prior distributions, the resulting posteriors follow a gamma distribution and therefore are all proper.

2.9.4 Choice of Prior Distribution

The choice of which prior distribution will give the best results is controversial and is often subjective. Four priors will be considered in this section and a simulation study based on the coverage rate is done to determine which prior will be ideal and produce the best results. The parameters of the conjugate prior distribution for the South African data are estimated from previous studies dating back from 2010 until 2013 for which, the mean and variance are calculated for each province and summed to yield the total mean and variance for 2010-2013. The parameters of the prior are then estimated by using the mean and variance of a gamma distribution and solving for the two parameters, α and β. This resulted in $\alpha = 13.78723$ and $\beta = 0.765957$. The data used to determine the values of α and β is listed in Table 2.5 and is based on the number of deaths due to acute pericarditis that took place in South Africa between 2010-2013.

The parameters for the Georgia oral cancer data are calculated based on approximately 30% of the data used in the study. The 30% of the data had a mean of 1.347826 and a variance of 5.78744 which resulted in parameters with values of $\alpha = 0.3139$ and $\beta = 0.2329$ for the Georgia data respectively, these parameters will be used in model 1 in Section 3.3.2.2.

Table 2.5: The observed number of deaths of acute pericarditis (Classification number: *I30*) in South Africa by province in 2010-2013.

Province	2010	2011	2012	2013	Total	Mean	Variance
Eastern Cape	1	1	2	2	6	1.5	0.3333
Free State	0	2	0	1	3	0.75	0.9167
Gauteng	3	6	3	3	15	3.75	2.25
KwaZulu-Natal	9	5	4	2	20	5	8.6667
Limpopo	4	2	0	1	7	1.75	2.9167
Mpumalanga	0	1	0	0	1	0.25	0.25
North-West	0	1	1	1	3	0.75	0.25
Northern Cape	0	0	0	3	3	0.75	2.25
Western Cape	1	6	2	5	14	3.5	5.6667
Total	**18**	**24**	**12**	**18**	**72**	**18**	**23.5**

The simulation algorithm is well-known and has the following steps:

1. For a given θ, simulate data from a Poisson distribution.
2. Using the data simulated in step 1, simulate θ from the posterior distribution.
3. Using the data simulated in step 1, repeat step 2 n times. In this case $n_T = 10000$.
4. Order the values obtained in step 3 in ascending order, such that $\theta_{(1)}, \theta_{(2)}, \theta_{(3)}, \ldots \theta_{(10000)}$.
5. The 95% credibility interval will be the values of $\theta_{(250)}, \theta_{(9750)}$.
6. Repeat steps 1 to 3, $n_T = 10000$, times then determine the number of credibility intervals which contain the true parameter θ.
7. Calculate the average interval length and standard deviation of the intervals as follows:

$$Average\ length = \frac{1}{n_I} \sum_{j=1}^{n_I} I_j$$

and

$$Standard\ Deviation = \sqrt{\frac{1}{n_I - 1} \sum_{j=1}^{n_I} \left(I_j - \bar{I}\right)^2},$$

where I_j is the interval length of the j^{th} interval and n_I is the number of intervals.

The results from the simulation study are given in the tables that follow and also include the average interval length and standard deviation. Code for this simulation can be found in the Appendix C.5. The main criteria for the selection of the prior distribution is the mean coverage rate followed by the mean

standard deviation and mean interval length. The mean coverage rate should be close to the theoretical value of 0.95 for a good prior distribution.

Table 2.6: The coverage rate, mean and standard deviation of the conjugate prior when $\alpha = 13.78723$, $\beta = 0.765957$ and $n_T = 10000$.

Value of θ	1	2	3	4	5	6	7	8	9	Mean
Coverage Rate	0.9491	0.9523	0.9526	0.9532	0.9505	0.9515	0.9472	0.9545	0.9514	0.9514
Mean Length	8.7620	9.3020	9.8110	10.2976	10.7625	11.2054	11.6361	12.0488	12.4504	10.4751
Standard Deviation	0.0904	0.0941	0.0976	0.1042	0.1091	0.1131	0.1182	0.1206	0.1252	0.1081

Table 2.6 shows the results of the simulation for the conjugate prior with the South African data parameters for α and β. The mean coverage rate is 0.9514 which is a difference of 0.0014 from the theoretical value of 0.95. The mean interval length is high at 10.4751 and the mean standard deviation is 0.1081.

Table 2.7: The coverage rate, mean and standard deviation of the conjugate prior when $\alpha = 0.3139$, $\beta = 0.2329$ and $n_T = 10000$.

Value of θ	1	2	3	4	5	6	7	8	9	Mean
Coverage Rate	0.9497	0.9483	0.9492	0.9493	0.9537	0.9506	0.9468	0.9496	0.9496	0.9496
Mean Length	4.0776	5.5413	7.6313	8.8781	9.514	10.9244	11.8121	12.389	13.127	9.3216
Standard Deviation	0.0620	0.0843	0.0943	0.1034	0.1104	0.1217	0.1287	0.1256	0.1305	0.1068

The parameters, α and β, of the Georgia data give the results shown in Table 2.7. The mean coverage rate is close to the theoretical value of 0.95 with a difference of only 0.0004. The mean interval length is high at 9.3216, a high mean interval length was also seen in the South African data.

Other values of α and β for the conjugate prior are also considered and are given in the following tables:

Table 2.8: The coverage rate, mean and standard deviation of the conjugate prior when $\alpha = 0.5$, $\beta = 6$ and $n_T = 10000$.

Value of θ	1	2	3	4	5	6	7	8	9	Mean
Coverage Rate	0.9514	0.9483	0.9508	0.9534	0.9530	0.9543	0.9497	0.9461	0.9479	0.9505
Mean Length	0.8123	1.1162	1.3692	1.5834	1.7703	1.9393	2.0947	2.2391	2.3755	1.7
Standard Deviation	0.0129	0.0152	0.0170	0.0186	0.0200	0.0213	0.0225	0.0239	0.0253	0.0196

The mean coverage rate for the simulation when using $\alpha = 0.5$ and $\beta = 6$ is closer to 0.95 than when the South African data parameters were used. The Georgia parameters have a difference of 0.0004 from the theoretical value of 0.95 which is less than the difference in the mean coverage rate in Table 2.8. The mean interval length, for $\alpha = 0.5$ and $\beta = 6$, is much smaller than the mean interval length for the South African and Georgia parameters.

Table 2.9: The coverage rate, mean and standard deviation of the conjugate prior when $\alpha = 3,\ \beta = 5$ and $n_T = 10000$.

Value of θ	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Mean**
Coverage Rate	0.9483	0.9525	0.9530	0.9512	0.9493	0.9478	0.9522	0.9516	0.9490	0.9505
Mean Length	1.4228	1.6880	1.9200	2.1280	2.3179	2.4945	2.6587	2.8139	2.9613	2.2672
Standard Deviation	0.0173	0.0196	0.0215	0.0236	0.0251	0.0267	0.0283	0.0291	0.0309	0.0247

Table 2.10: The coverage rate, mean and standard deviation of the conjugate prior when $\alpha = 8,\ \beta = 0.25$ and $n_T = 10000$.

Value of θ	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Mean**
Coverage Rate	0.9537	0.9469	0.9471	0.9490	0.9495	0.9494	0.9513	0.9499	0.9506	0.9497
Mean Length	9.8099	10.7483	11.6158	12.4221	13.1857	13.9066	14.5905	15.2428	15.8745	13.0440
Standard Deviation	0.1047	0.1151	0.1224	0.1289	0.1375	0.1458	0.1507	0.1570	0.1607	0.1359

The results from the conjugate prior simulation show that the South African data parameters gave the poorest coverage rate out of all the conjugate priors. The South African data parameters, therefore will not be used as the parameters of the hyperprior distributions. The mean interval length of the simulation when using the parameters $\alpha = 8$ and $\beta = 0.25$ has the highest interval length of all the conjugate priors and thus will not be used. The Georgia data parameters give the closest mean coverage rate to the theoretical value of 0.95 than all of the other parameters and thus the Georgia parameters will be used in the hyperprior distribution.

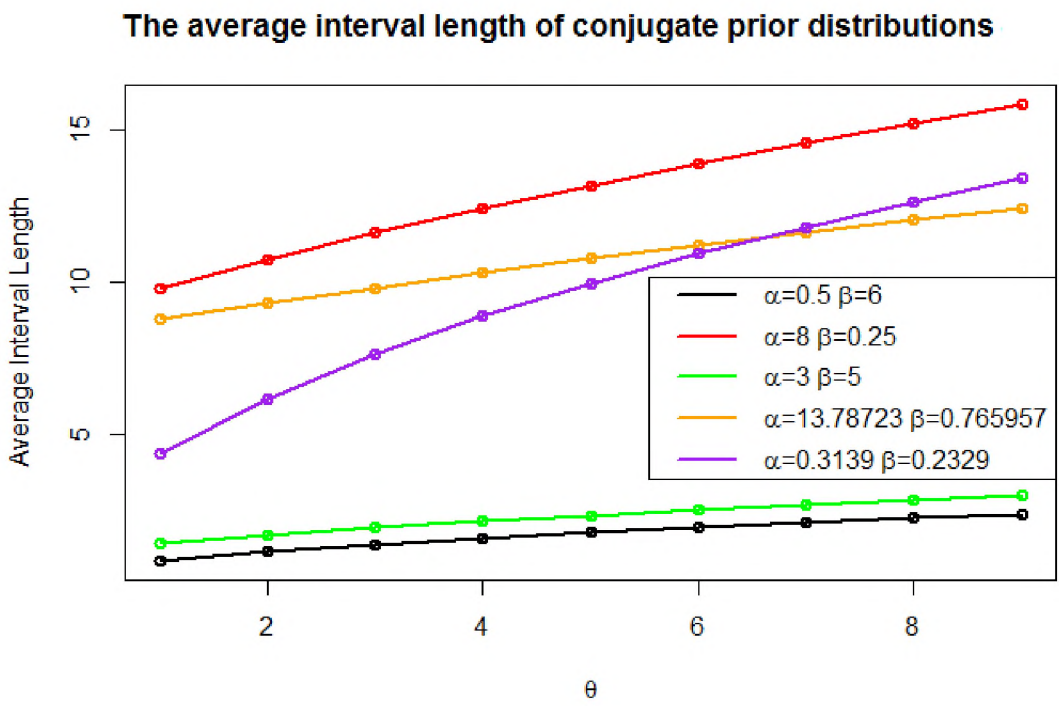


Figure 2.11: The average interval length of the conjugate prior distributions with different values of α, β and θ.

The prior distributions are chosen based on the smallest average interval length, after first examining the average coverage rate and average standard deviation. The prior distributions with parameters $\alpha = 0.5,\ \beta = 6$ and $\alpha = 3,\ \beta = 5$ show the smallest average interval lengths in Figure 2.11, although the coverage rates of these two prior distributions are not better than some of the other prior distributions considered in this simulation study.

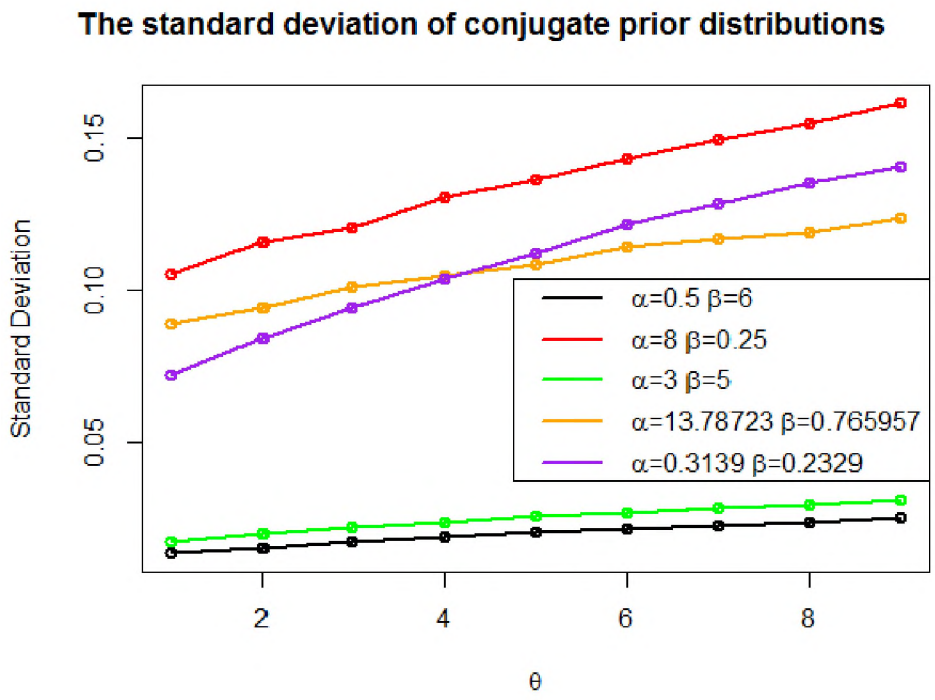


Figure 2.12: The standard deviation of the conjugate prior distributions with different values of α, β and θ.

The prior distributions with the lowest standard deviations are considered the best prior distributions after first considering the average coverage rate. Figure 2.12 shows the standard deviation for the conjugate prior distributions in the study. The prior distributions with parameters $\alpha = 0.5,\ \beta = 6$ and $\alpha = 3,\ \beta = 5$ show the lowest standard deviations.

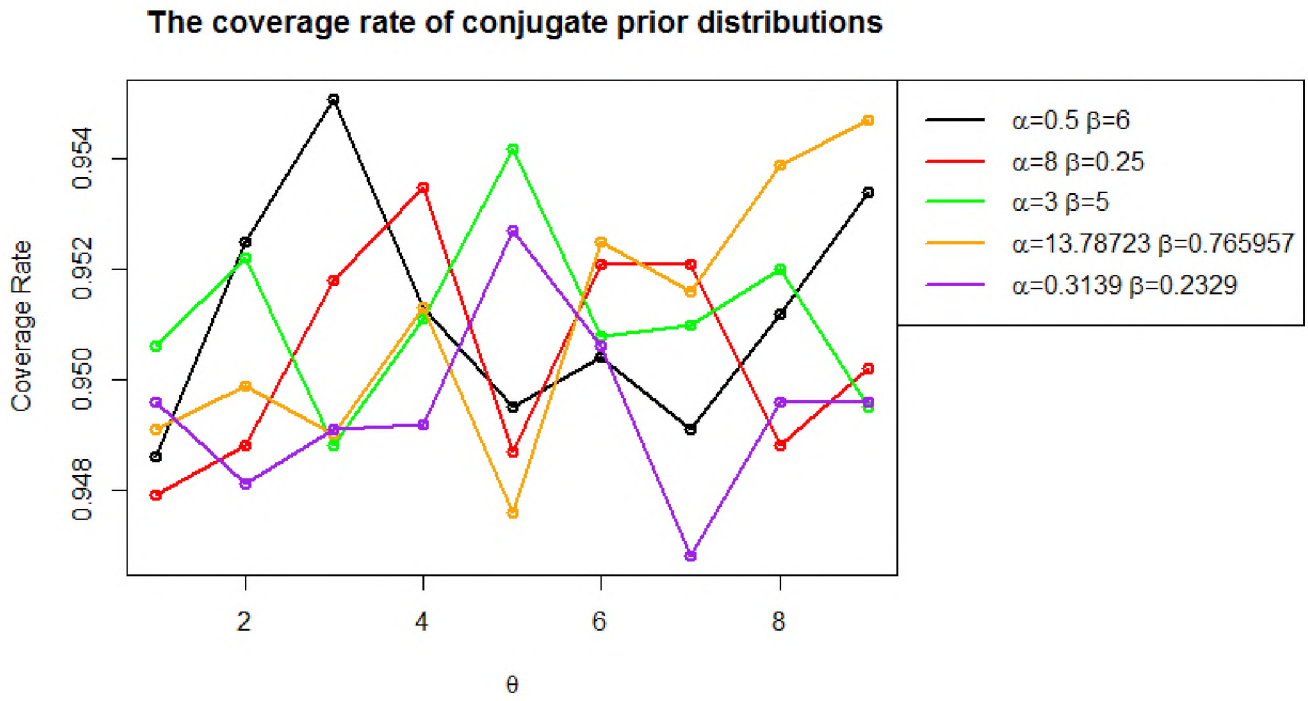


Figure 2.13: The coverage rate of the conjugate prior distributions with different values of α, β and θ.

Figure 2.13 shows the coverage rate for the conjugate prior distributions in the simulation study. It is difficult to make conclusions based on the figure as the coverage rate changes with the value of θ. The average coverage rate is thus used. It may also be possible to consider the most constant coverage rate and then consider the bias associated with the different prior distributions as an alternative to the average coverage rate.

Table 2.11: The coverage rate, mean and standard deviation of the Jeffreys' prior with $n_T = 10000$.

Value of θ	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Mean**
Coverage Rate	0.9489	0.9528	0.9474	0.9509	0.9536	0.9508	0.9474	0.9480	0.9480	0.9498
Mean Length	4.2922	5.7344	6.8974	7.9176	8.8441	9.6819	10.4208	11.1544	11.7925	8.5261
Standard Deviation	1.5973	1.8482	1.9099	1.9458	1.9417	1.9376	1.9624	1.9842	2.0183	1.9050

The mean coverage rate for the Jeffreys' prior is 0.9498, a slight difference of 0.0002 from the theoretical value of 0.95. This is the closest mean coverage rate to the theoretical value of 0.95 out of all of the prior distributions considered in this thesis. The mean of the average interval length is 8.5261 compared to 8.5295 and 8.7764 of the probability matching prior and divergence prior, respectively. The Jeffreys' prior and probability matching prior should produce almost the same results as the posterior distribution of both prior distributions was shown to be the same in Section 2.9.3. There are slight differences between the results of the Jeffreys' prior and probability matching prior due to the randomness in the simulation.

Table 2.12: The coverage rate, mean and standard deviation of the probability matching prior with $n_T = 10000$.

Value of θ	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Mean**
Coverage Rate	0.9515	0.9523	0.9521	0.9533	0.9523	0.9497	0.9521	0.9541	0.9511	0.9521
Mean Length	4.3056	5.7349	6.9308	7.9156	8.8248	9.6969	10.4216	11.1379	11.7978	8.5295
Standard Deviation	1.5965	1.8411	1.9203	1.9391	1.9599	1.9636	1.9568	1.9391	1.9820	1.8998

The results of the simulation of the probability matching prior are given in Table 2.12. The results show that the Jeffreys' prior has slightly better results as the mean coverage rate is closer to the theoretical value of 95% and mean length of the interval is lower than that of the probability matching prior although the mean standard deviation is slightly higher in the results of the Jeffreys' prior than in the results of the probability matching prior.

Table 2.13: The coverage rate, mean and standard deviation of the divergence prior with $n_T = 10000$.

Value of θ	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Mean**
Coverage Rate	0.9546	0.9507	0.9518	0.9503	0.9502	0.9527	0.9516	0.9514	0.9470	0.9511
Mean Length	4.7439	6.0651	7.1926	8.1571	9.0602	9.8545	10.6342	11.3377	11.9419	8.7764
Standard Deviation	1.4699	1.7355	1.8298	1.8847	1.9308	1.9234	1.9447	1.9250	1.9350	1.8421

A comparison of the results of the divergence prior in Table 2.13 and the results of the Jeffreys' prior in Table 2.11 shows that the Jeffreys' prior produces better results than the divergence prior. The mean coverage rate is closer to the theoretical value of 95% and mean interval length of the Jeffreys' prior is lower than that of the divergence prior. The mean of the standard deviation of the divergence prior is, however, lower than the mean of the standard deviation of the Jeffreys' prior. This is a slight difference of only 0.0629.

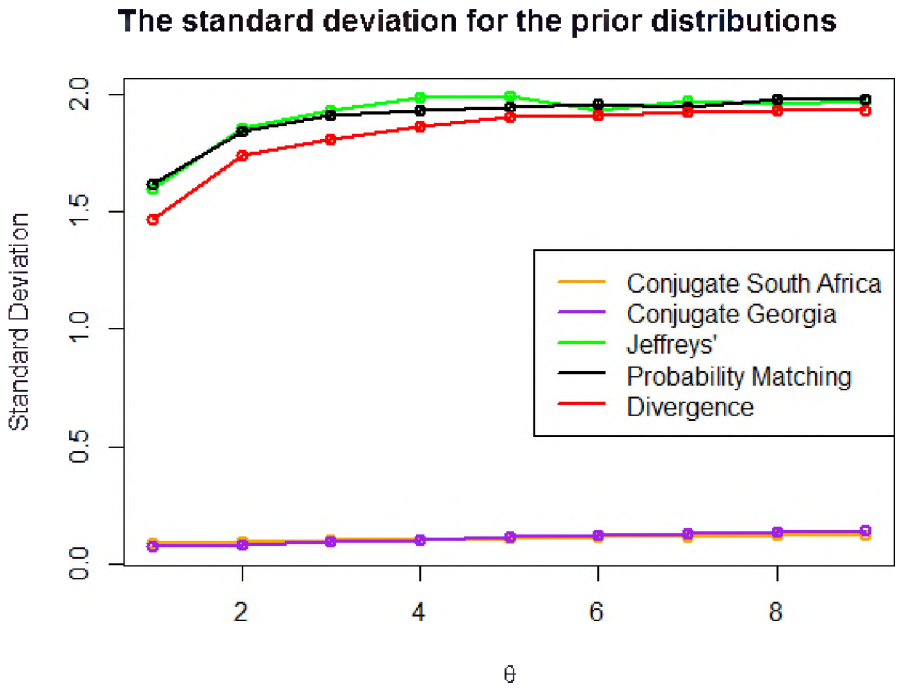


Figure 2.14: The standard deviation of the prior distributions.

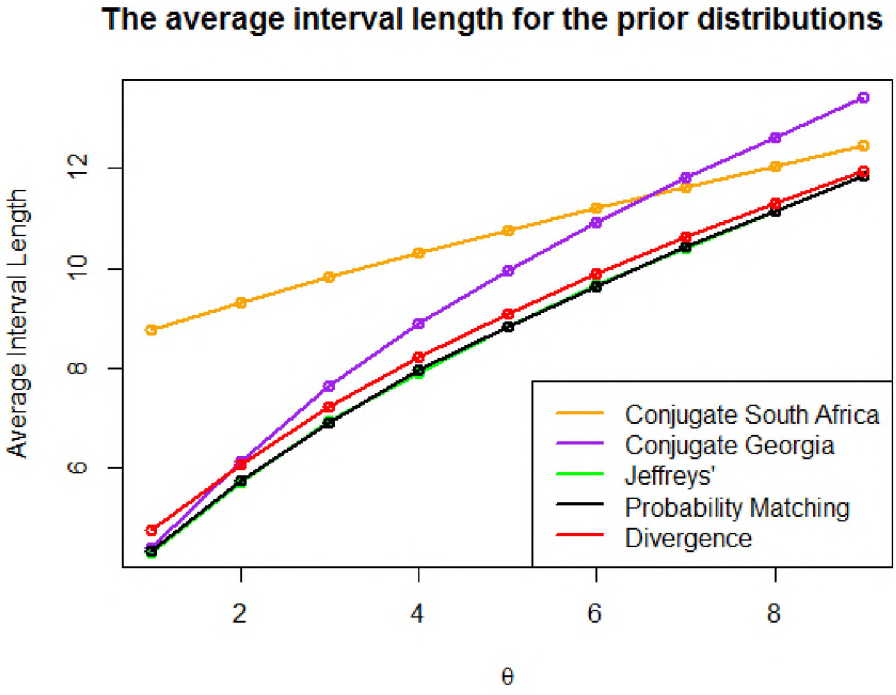


Figure 2.15: The average interval length of the prior distributions.

The average interval length increases as the value of θ increases, this is evident in Figure 2.15.

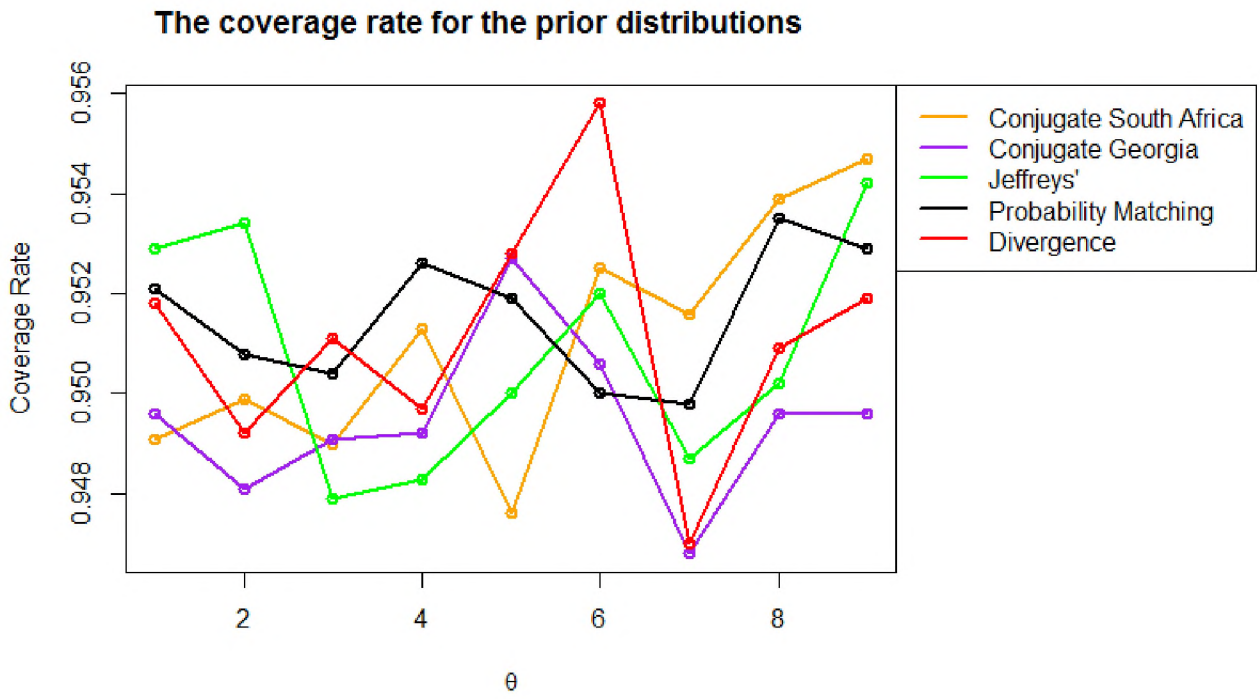


Figure 2.16: The coverage rate of the prior distributions.

The simulation results show that Jeffreys' prior has the highest mean coverage rate and thus is the best prior of all the priors considered in the simulation. The Jeffreys' prior will thus be used in the South African and the Georgia models. The Georgia models will also use the conjugate prior with the Georgia data parameters, as this was the second best prior in the simulation. The last Georgia model will have a fixed parameter for the precision of the uncorrelated heterogeneity (UH) component.

2.10 Markov Chain Monte Carlo (MCMC)

Gelman et al. (2014) define MCMC as a method whose fundamental property is to draw samples of θ from a proposal distribution and then sequentially drawing samples closer and closer to the target posterior distribution, denoted by $\pi(\theta|y)$. The defining property of an MCMC is that the distribution of the sampling values depends only on the last value sampled. A more formal definition is provided by Gelman et al. (2014) which says that a Markov chain is a sequence of random variables, denoted by $\theta^1, \theta^2, \ldots, \theta^k$, such that for any k, the distribution of θ^k depends only on the previous value of the chain θ^{k-1} given the preceding history of the chain. The MCMC method is widely used in Bayesian statistics.

2.10.1 The Gibbs Sampling Algorithm

2.10.1.1 Overview

The Gibbs sampler was first used by Geman and Geman in 1984 who applied the algorithm in image processing, according to Casella & George (1992). Earlier versions of the algorithm appeared in at least 1953 and it was eventually adapted by Hastings in 1970. Lesaffre & Lawson (2012) state that the Gibbs distribution was applied to the intensity of the pixels in an image. The distribution can have millions of unknowns and hence Geman and Geman, developed in 1984, a sampling algorithm to enable inference from the distribution. The Gibbs Sampling Algorithm only became popular in the 1990's when Gelfand and Smith showed its potential in statistical applications, and this was one of the reasons for the revival of Bayesian statistics as explained by Casella & George (1992). The main condition for the use of the Gibbs sampler is that one must be able to specify and sample from all the conditional distributions of the parameter, that is for a parameter $\boldsymbol{\theta}$ and data $\mathbf{y} = \Big\{y_1, y_2, \ldots, y_N\Big\}$, one must be able to specify and sample from:

$$h\Big\{(\theta_1|\theta_2, \theta_3, \ldots, \theta_p,\ y);\ (\theta_2|\theta_1, \theta_3, \ldots, \theta_p,\ y);\ \ldots (\theta_p|\theta_1, \theta_2, \ldots, \theta_{p-1},\ y)\Big\}.$$

2.10.1.2 The bivariate Gibbs sampler

Lesaffre & Lawson (2012) define the bivariate Gibbs sampler as follows:

Suppose the joint distribution of the parameters θ_1 and θ_2 is given by: $\pi\Big(\theta_1, \theta_2|\mathbf{y}\Big)$ with data $\mathbf{y} = \Big\{y_1, y_2, \ldots, y_N\Big\}$ and suppose that the joint distribution is uniquely determined by all its conditional distributions, such that it is determined by: $\pi\Big(\theta_1|\theta_2, \mathbf{y}\Big)$ and $\pi\Big(\theta_2|\theta_1, \mathbf{y}\Big)$. We want to obtain a sample from the joint distribution $\pi\Big(\theta_1, \theta_2|\mathbf{y}\Big)$ but as this is a complex distribution we must first obtain starting values which may be estimates or initial guesses. For this case only one starting value is needed, suppose we use two starting values and we denote them by θ_1^0 and θ_2^0, respectively. The sampler then samples to obtain θ_1^k and θ_2^k where $k = 1, 2, \ldots$. Thus each iteration of the Gibbs sampler cycles through the sub-vectors of $\boldsymbol{\theta}$, drawing each subset conditioned on all the other values. The sampler forms $\theta_1^1, \theta_2^1, \theta_1^2, \theta_2^2, \theta_1^3, \theta_2^3, \ldots$ which is a chain of dependent values where θ^k is only dependent on the previous value θ^{k-1} such that θ^k is independent of the entire preceding history of the chain. The sampled values are taken as draws from the posterior distribution $\pi\Big(\theta_1, \theta_2|\mathbf{y}\Big)$ after a burn-in period denoted by N_B. The initial part of the chain is discarded, there are a number of methods to determine the number of iterations before the burn-in period. The algorithm is given in Lesaffre & Lawson (2012) as:

- Starting values θ_1^0, θ_2^0
- Sample $\theta_1^{(1)}$ from $\pi\Big(\theta_1|\theta_2^0, \mathbf{y}\Big)$

- Sample $\theta_2^{(1)}$ from $\pi\left(\theta_2|\theta_1^{(1)}, \mathbf{y}\right)$

$\vdots$

$\longrightarrow N_B \Rightarrow$Burn-in period

- Sample $\theta_1^{(k)}$ from $\pi\left(\theta_1|\theta_2^{(k-1)}, \mathbf{y}\right)$
- Sample $\theta_2^{(k)}$ from $\pi\left(\theta_2|\theta_1^{(k-1)}, \mathbf{y}\right)$

2.10.1.3 The general Gibbs sampler

A general algorithm for the Gibbs sampler is provided in Lesaffre & Lawson (2012) and is given as:

Starting with initial values $\boldsymbol{\theta}^0 = \left(\theta_1^0, \theta_2^0, \ldots, \theta_d^0\right)$ the algorithm performs these d steps at iteration $(k+1)$:

- Sample $\theta_1^{(k+1)}$ from $\pi\left(\theta_1|\theta_2^k, \ldots, \theta_{(d-1)}^k, \theta_d^k, \mathbf{y}\right)$
- Sample $\theta_2^{(k+1)}$ from $\pi\left(\theta_2|\theta_1^{(k+1)}, \theta_3^k, \ldots, \theta_d^k, \mathbf{y}\right)$

$\vdots$

- Sample $\theta_d^{(k+1)}$ from $\pi\left(\theta_d|\theta_1^{(k+1)}, \theta_2^{(k+1)}, \ldots, \theta_{(d-1)}^{(k+1)}, \mathbf{y}\right)$

2.10.1.4 Advantages of the Gibbs sampler

The advantages of the Gibbs sampler are given in Lawson (2009) as:

- A single new θ value is computed at each iteration.
- When the conditional distributions are simple, the chain may converge faster.
- The sampler is often used in simpler hierarchical modelling applications.
- The implementation of the algorithm is straightforward.
- It can be extended to other sampling applications, for example, reversible Gibbs sampler and Random-scan Gibbs sampler.
- The Gibbs sampler can be combined with other algorithms to sample from full conditionals.
- The Gibbs sampler is used in many statistical programs.
- The Gibbs sampler can simplify complex problems by converting high-dimensional problems into lower-dimensional problems.

2.10.1.5 Disadvantages of the Gibbs sampler

Lawson (2009) provides the following disadvantages:

- One may be unable to specify all of the conditional distributions.
- Derivation of the conditional distributions may be time-consuming.
- Block updates of parameters may not be available in Gibbs sampling.
- The Gibbs sampler is not often used in complex problems.
- The Gibbs sampler may not converge.
- The posterior distribution must be known.

2.10.2 The Metropolis-Hastings Algorithm

2.10.2.1 Overview

The Metropolis-Hastings algorithm was developed by Metropolis in 1953 with an intended purpose in mechanical physics as explained by Lesaffre & Lawson (2012). The algorithm was further developed by Hastings in 1970 with the application being in a statistical sense. The algorithm remained virtually unused as the processing power and efficiency of computers was very poor in the 1970's. With the increase in processing power and efficiency of computers in the 1990's and the re-emergence of Bayesian statistics, the algorithm became more and more popular. The algorithm can be used to simulate from any distribution as long as the analytical form of the distribution is known. This is in contrast to the Gibbs sampler which requires the posterior distribution and the conditional distributions to be known. The Gibbs sampler is a special case of the Metropolis-Hastings Algorithm.

2.10.2.2 The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm needs a target distribution, denoted by π, to be defined before the algorithm is executed. A proposal density, denoted by $q(y|x)$, is defined in Robert & Casella (1999) as taken in terms of the ruling quantity in the model. The algorithm works well when the proposal density is symmetric $q(y|x) = q(x|y)$, then the probability with which the new value, which is generated by the algorithm, is accepted is the ratio probability of the new value and the previous value. When the new value has a higher probability than the current value, the new value is chosen and these values when put together then make up a sample from π. The proposal density also works well when $\frac{\pi(y)}{q(y|x)}$ can be determined up to a constant value which is independent of the value of x, as written by Robert & Casella (1999).

Let $x^{(0)}$ be an arbitrary initial value and given the value of $x^{(t)}$, where $t = 1, 2, \ldots$, then the Metropolis-Hastings algorithm is executed by Robert & Casella (1999) in the following steps:

1.
$$\text{Simulate } Y_t \sim q\Big(y|x^{(t)}\Big)$$

2.
$$\text{Let } \rho(x,y) = min\left\{\frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}, 1\right\}$$

3.
$$\text{Sample } \; X^{(t+1)} = \begin{cases} Y_t & \text{with probability } \rho\Big(x^{(t)}, Y_t\Big) \\ x^{(t)} & \text{otherwise} \end{cases}$$

An advantage of using the Metropolis-Hastings algorithm is that the values of y_t may be used such that the ratio of:

$$\frac{\pi\Big(y_t\Big)}{\pi\Big(x^{(t)}\Big)}$$

may be decreased. Metropolis-Hastings algorithm only depends on:

$$\frac{\pi\Big(y_t\Big)}{\pi\Big(x^{(t)}\Big)},$$

and

$$\frac{q\Big(x^{(t)}|y_t\Big)}{q\Big(y_t|x^{(t)}\Big)},$$

as written by Robert & Casella (1999).

There are many methods available to construct a proposal density, one of which is the random-walk proposal which is written as: $X_{n+1} = X_n + \epsilon_n$ where ϵ_n is distributed as a symmetric random variable around 0 and is simulated independent from $X_n,\ X_{n-1},\ X_{n-2}, ...$

One must also consider the variance of the proposal density as the size of the variance affects the acceptance ratio of the chain. If the variance is too small, the chain will have a high acceptance ratio and the chain will move slowly over the sample, reducing efficiency. If the variance is too large many proposal values will be rejected and the chain will stay in one place for a long time, which also reduces the efficiency of the chain. Therefore the variance of the proposal density must be chosen carefully to ensure a good acceptance ratio and a high chain efficiency.

2.10.2.3 Advantages of the Metropolis-Hastings Algorithm

According to Lawson (2009), the advantages include:

- The algorithm does not require full conditional distributions.

- There are an infinite number of proposal densities which result in a Markov chain that converges to the target density.
- The algorithm can simulate from any distribution as long as the analytical form is known.
- Enables block updates for a parameter.
- The algorithm is often used in complex problems.

2.10.2.4 Disadvantages of the Metropolis-Hastings Algorithm

According to Lawson (2009), the disadvantages include:

- Selection of the proposal density may be difficult.
- The convergence of the algorithm is slower than the Gibbs sampler.
- It is more difficult to implement than the Gibbs sampler.
- The algorithm may not converge.
- The algorithm does not guarantee the new value will be accepted.

2.11 Convergence and Diagnostic Tests

2.11.1 Overview

For the MCMC method, convergence needs to be determined up to a number of iterations. The purpose of convergence, proposed by Lawson (2009), is to determine whether the distribution of the sampled values reaches the equilibrium distribution of the Markov chain. A burn-in period, which varies between experiments, also needs to be determined. The burn-in period is the number of iterations needed to ensure independent and accurate samples are provided. A poor burn-in period results in a Markov chain getting stuck at one point. Lesaffre & Lawson (2012) reiterate that convergence is not easily determined and convergence needs to be determined to ensure the posterior distribution is represented accurately. Convergence may be easily determined in simple problems but in more complex problems graphical and diagnostic tests are required. Lawson (2009) says that determining convergence is considered an art rather than a science.

2.11.2 Convergence of an Algorithm

Brooks et al. (2003) state that because it is difficult to determine the number of iterations and since a parameters state might change between iterations, one needs to choose a parameter which uniquely defines the model. It is easier to monitor the convergence of this single parameter than the convergence of the entire model. This is more commonly known as determining convergence for a marginal distribution. Lesaffre & Lawson (2012) define convergence as an "asymptotic property" of a Markov chain for which the distribution of θ^k, denoted by $\pi_k(\theta)$, tends to the equilibrium distribution as $k \to \infty$. Brooks et al. (2003) explain that another way to detect convergence is to let as many parameters as possible keep their interpretation as they are sampled from one model to another. This will allow these parameters to be checked and then convergence can also be checked. This process uses the " analysis of variance (ANOVA)-type decomposition" output.

Lesaffre & Lawson (2012) name two elements to check for convergence, these are (1) monitor the stationarity of the Markov chain and (2) check the accuracy of the posterior summary measures. The stationarity element involves determining the burn-in part of the algorithm, that is to determine the iteration, n_0, such that $k \geq n_0$, then θ^k may be sampled from the posterior distribution. The accuracy element ensures that the posterior summary measures are sampled with a predetermined degree of accuracy.

2.11.3 Techniques to Determine Convergence

Many techniques may be used to determine convergence. The more popular techniques are the ones which are built into statistical software, and according to Lesaffre & Lawson (2012), many techniques are very complex and difficult to analyse thus are not used often. Experimenters may use a combination of techniques because each technique has its own weaknesses.

Cowles & Carlin (1996) propose two areas to focus on when trying to determine convergence. Firstly, the number of iterations are predetermined by looking at the Markov transition kernel and secondly, is using diagnostic tools to analyse the output of the Markov chain. Lesaffre & Lawson (2012) separate the techniques into two categories, namely graphical techniques and diagnostic tests.

2.11.3.1 Graphical Techniques

2.11.3.1.1 Trace Plot A trace plot is one of the most simple and important graphical techniques. Lesaffre & Lawson (2012) use the trace plot to monitor the Markov chain univariately by plotting each parameter on its own plot. The trace plot is drawn by evaluating the lag of the likelihood. The main purpose of the trace plot is to show the chain mixing rate and convergence of the model.

2.11.3.1.2 Autocorrelation Plot The autocorrelation plot is used to show the correlation of lags. Lesaffre & Lawson (2012) use the autocorrelation plot to show the mixing rate of the chain.

2.11.3.1.3 Cross-Correlation Plot The cross-correlation plot is used to show model over-specification. It is used as a measure of the correlation between variables.

2.11.3.2 Diagnostic Tests

2.11.3.2.1 Brooks-Gelman-Rubin (BGR) Diagnostic The BGR diagnostic is used when the posterior is multi-modal according to Lesaffre & Lawson (2012). The steps of the BGR diagnostic are described in Cowles & Carlin (1996) as:

1. Before the algorithm is implemented, an overdispersed estimate of the target distribution is obtained. The initial values are then generated from the overdispersed estimate.
2. After the Gibbs Sampler has been executed for n iterations. The last $\frac{n}{2}$ iterations are used to obtain another target distribution of the scalar quantity as a student's t-distribution.

Disadvantages of the BGR diagnostic include:

- The initial values of the overdispersed estimate may be difficult to find. This step needs an experienced user.
- Normal approximation for analysing convergence may be unreliable.

2.11.3.2.2 Geweke Diagnostic Lesaffre & Lawson (2012) suggest that the Geweke diagnostic tests the stationarity of a Markov chain by analysing the means of the first part of the chain to the means of the last part of the chain. This method creates a space between the two means which implies that they can be treated as independent. Let the number of iterations in the chain be denoted by n_T and the number of iterations in the first and last part of the chain be denoted by n_A and n_B, respectively. Let $\bar{\theta}_A$ and $\bar{\theta}_B$ denote each mean. Then a frequentist significance test can be used to test the stationarity of the chain:

$$Z = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}.$$

Lesaffre & Lawson (2012) explain that since the elements of a Markov chain are dependent, $\frac{S_A^2}{n_A}$ and $\frac{S_B^2}{n_B}$ underestimate the variances and the means are dependent. Cowles & Carlin (1996) introduce the assumption that the nature of the MCMC algorithm and g, where g are the output values of the algorithm, imply a spectral density denoted by $S_g(w)$. When this assumption is made then $E\Big(g(\theta)\Big)$ can be estimated by:

$$g_{n_T} = \frac{\sum_{i=1}^{n_T} g\Big(\theta^{(i)}\Big)}{n_T},$$

for n iterations and the asymptotic variance is given by Cowles & Carlin (1996) as $\frac{S_g(0)}{n_T}$.

2.11.4 Improving Convergence

There are many methods and techniques available to improve convergence but since every problem is unique, some methods and techniques improve some problems while others have no effect. The type of application of the problem will determine which techniques and methods will work the best. The methods and techniques to improve convergence are separated into the categories: Burn-in period, Thinning, Choice of initial values, Transformation, Reparameterisation and the number of chains.

2.11.4.1 Burn-in Period

Brooks (1998) explains that the number of iterations in the burn-in period is important because the burn-in period is designed to limit the "inferential bias" created by the starting values. The burn-in period changes between different problems and hence estimation of the burn-in period is a common problem. When "geometric ergodicity" is valid, Brooks (1998) defines the transition density, denoted by $\pi'(x, \cdot)$, which has k steps as:

$$|\pi'(x, \cdot) - \pi(\cdot)| \leq M(x)\rho',$$

where $\rho \, \epsilon \, \mathbb{R}$ and for some value of M. The algorithm may be stopped when $|\pi'(x, \cdot) - \pi(\cdot)| \leq \epsilon$, when $\epsilon > 0$ and then the burn-in period is estimated by:

$$N_B^* = \frac{log\left\{\epsilon / M(x)\right\}}{log(\rho)}, \text{ where } \rho \, \varepsilon \, \mathbb{R}, \; k > 0.$$

This, however, is seldom used as it is challenging to show that there is a "geometric rate of convergence to stationarity."

Diaconis & Stroock (1991) introduced a method for determining the bounds for the second largest eigenvalue and the spectral gap in a reversible Markov chain. The method uses the discrete version of the poincaré inequality and according to Brooks (1998), the method is used to estimate the flow rate of the chain between states to bound a convergence rate.

Edwards & Sokal (1988) generalise the Monte Carlo algorithm to arbitrary models and essentially provides a way to estimate the convergence rate of an arbitrary model based on Cheeger's inequality, according to Brooks (1998).

2.11.4.2 Thinning

Brooks (1998) states that since each problem is different, the number of iterations that the Markov chain runs for varies from problem to problem. The computation time and processing power of computers may be a limiting factor in choosing the number of iterations. The method of thinning is then used to reduce the number of iterations. Lesaffre & Lawson (2012) define thinning as a method to lower

the autocorrelation by only saving every k^{th} value of the chain. This method, for lags that are greater than one, may minimise autocorrelations until they are all zero. Lesaffre & Lawson (2012) state that the new Markov chain where every k^{th} value has been saved has a higher Monte Carlo error than the original chain.

Brooks (1998) proposes the method to estimate the sample size of an IID sample. Since the standard deviation is given by $\frac{\sigma}{\sqrt{n}}$, where σ is the standard deviation of the posterior distribution of θ, and if an estimate of σ is available then the approximate sample size can be calculated.

2.11.4.3 Choice of Initial Values

The initial values will be independent of the MCMC output after the burn-in period. Brooks (1998) explains that the choice of initial values will affect the speed and performance of the chain.

Lesaffre & Lawson (2012) introduce the idea that the choice of the initial values may affect the mixing rate of the Markov chain. The mixing rate may be low when the initial values are chosen such that the posterior probability is closer to zero. The chain may then get 'stuck' in an area for a long period of time. The technique to identify this problem involves using a trace plot. The trace plot will have an increasing or decreasing line when the initial values are chosen such that the posterior probability is close to zero in that area. When more than one chain is run, Gelman & Rubin (1992) proposed that the distribution of the initial values must be overdispersed with respect to the target distribution, in order to detect convergence accurately.

Methods for selecting initial values are listed in Brooks (1998) as: assigning hyperparameters to fixed values, removing missing data and checking to see if the initial values may be estimated from the maximum likelihood. Gelman & Rubin (1992) suggest a mode-finding algorithm to find areas where the density is high. The initial values are then generated by sampling from t-distributions at these areas.

2.11.4.4 Reparameterisation

Reparameterisation is used to reduce the correlation between variables in a Markov chain, according to Brooks (1998). High correlation between variables changes the way in which the MCMC algorithm runs. The algorithm has a slower convergence and a higher computation time when there is correlation between the variables. Lesaffre & Lawson (2012) say that reparameterisation may remove constraints on parameters. Reparameterisation in a linear regression is done by centring the regressors and is written in Lesaffre & Lawson (2012) as:

$$y = \beta_0 + \beta_1 x + \epsilon,$$

then $\theta = \left(\beta_0, \beta_1, \sigma\right)'$ changes to $\theta^* = \left(\beta_0^*, \beta_1, \sigma\right)'$ where $\beta_0^* = \beta_0 + \beta_1 \bar{x}$.

Brooks (1998) introduces approximate orthogonalisation which is used to obtain posterior parameters which are uncorrelated. This method, however, is only used in simple problems and it is not

appropriate for more complex problems. Another method proposed by Gelfand et al. (1995) is hierarchical centring which may be used in more complex problems but it can only be used when a model has a linear structure.

2.11.4.5 Transformations

Transformations are used when there is multicollinearity in the regressors, according to Lesaffre & Lawson (2012). Centring of the regressors is achieved by dividing each value of the regressor by its corresponding standard deviation. This method reduces the multicollinearity of the regressors.

2.11.4.6 The Number of Chains

The decision to run a Markov chain with more than one chain has its advantages and disadvantages. When one long chain is used, according to Brooks (1998), the result is a chain which will approximate the target distribution much better than any number of smaller chains would. When using more than one chain, the number of iterations that are removed for the burn-in period is more than that of one chain; this leads to additional computation time. The exploration of a sample is important and according to Brooks (1998), many chains explore the sample better and faster than a single chain does. Brooks (1998) proposes the regenerative method as an alternative to using many chains. This method uses a single chain which is restarted at "regeneration times", to yield many replications which are closer to the target distribution than any number of independent chains are.

2.12 Autocovariate Models

The autocovariate models incorporate spatial autocorrelation by predicting whether the response variable at one region replicates any response variable of a nearby region. Dormann et al. (2007) propose that this is modelled by incorporating a "distance-weighted" function of nearby response variables into the explanatory variables of the model and is named the autocovariate. The autocovariate can model spatial correlation from "conspecific attraction, limited dispersal, contagious population growth, and movement of censored individuals between sampling sites" as written by Dormann et al. (2007). The model which does not include the autocovariate is represented as:

$$y = X\beta + \epsilon,$$

including the autocovariate changes the model to:

$$y = X\beta + \phi Q + \epsilon,$$

where ϕ is the coefficient of the autocovariate Q. The autocovariate at any region i is calculated by Dormann et al. (2007) as:

$$Q_i = \sum_{j \varepsilon k_i}^{N} w_{ij} y_j \tag{2.15}$$

$$Q_i = \frac{\sum_{j \varepsilon k_i}^{N} w_{ij} y_j}{\sum_{j \varepsilon k_i}^{N} w_{ij}}, \tag{2.16}$$

where Equation 2.15 is the weighted sum and Equation 2.16 is the weighted average and y_j is the observed value at region j surrounded by the group of k_i neighbours and w_{ij} is the weight of region $j's$ impact on region i.

2.12.1 Random and Spatial Effects

The natural logarithm of the hierarchical model enables a spatial and random effects term to be included in the model as in DiMaggio (2012). The random effects in the model are said to cover the possible "group-level heterogeneity". Random effects allow for the presence of spatial autocorrelation and spatial heterogeneity in the model. Osei (2010) proposes that assigning a prior distribution for spatial effects creates a "spatial dependency structure" which considers the nature of the distribution of the regions in the area under study. There are two types of random effects named in Osei (2010) as additive and multiplicative random effects.

Overdispersion is very common in count data as mentioned in Section 2.8.1.2 and may be solved by using mixture models. Lawson (2009) recommends introducing a prior distribution modelled for the relative risk or introducing a random effect into the predictor term, both of which account for overdispersion. The Poisson-gamma model assigns a prior distribution to the relative risk and will be considered in later sections. Another recommendation to account for overdispersion is introduced by Osei (2010) who models the parameter λ of the Poisson distribution by introducing a parameter ϵ, which is an unobserved random variable to the model by either addition or multiplication. The first method being additive random effects and the second being multiplicative random effects. A conditional autoregressive (CAR) term which accounts for spatial effects is also included in the model. The CAR is derived by a group of "spatial neighbourhoods", where each neighbourhood contains nearby spatial observations which are in close proximity. An example of a model is given in DiMaggio

(2012) as:

$$\begin{aligned} y_i &\sim pois\Big(\mu_i\Big) \\ log\Big(\mu_i\Big) &= B_n + T_1 + T_2, \end{aligned}$$

where T_1 is the random effects term, T_2 is the spatial effects term and B_n is the log-linear terms which contain confounders. The CAR model makes use of additive random effects to allow for spatial heterogeneity.

2.12.2 Log-Normal Model

An extension of the canonical link function in Section 2.8 is to include a UH term, the link function then becomes:

$$\eta_i = \big(x_i'\beta + v_i'\gamma\big),$$

where v_i is the uncorrelated random effect as introduced by Lawson (2009). The UH is assigned a zero mean Gaussian prior distribution as proposed by Spiegelhalter et al. (2003) and is modelled as:

$$v_i \sim N(0, \phi).$$

Lawson (2009) provides two disadvantages of using the UH model with a relative risk modelled by a gamma prior distribution, these are:

The gamma distribution does not have an easy and flexible way of generalising the distribution to include spatially correlated parameters and "covariate adjustments" are difficult to model using a gamma distribution. Thus Lawson (2009) incorporates a log normal distribution to account for the additional variation in the model when the random effects are correlated. These are incorporated via additive random effects. A model with both UH and correlated heterogeneity (CH) may add more flexibility, Besag et al. (1991) propose that this can be modelled as:

$$\eta_i = exp\Big\{x_i'\beta + S_i + v_i\Big\},$$

where S_i is the CH component, v_i is as specified earlier and x_i are the explanatory variables. The prior distributions of S_i and v_i can be modelled by a variety of methods, although the Improper CAR (ICAR) and Proper CAR (PCAR) models will be considered for this research.

2.12.3 Conditional Autoregressive (CAR) Models

Suppose that there exists a set of spatially correlated Gaussian random effects, denoted by $S_1, S_2, \ldots, S_N$, for the N regions in the area under the study and suppose that the joint distribution of the random effects are:

$$\mathbf{S} \sim MVN\big(\boldsymbol{\mu}, \omega\Sigma\big),$$

where $\mathbf{S}$ is a vector containing $S_i,\ i = 1, 2, \ldots, N$ and is distributed as a multivariate normally distributed random variable of N-dimension, $\boldsymbol{\mu}$ is a mean vector of $1 \times N$ dimension, ω is a strictly non-negative value which governs the variability of S_i and Σ is a $N \times N$ positive definite matrix as in Spiegelhalter et al. (2003). The "between-region" covariance matrix $\omega\Sigma$ is expressed in Osei (2010) as:

$$\omega\Sigma = \omega\big(I - \rho W\big)^{-1} M, \tag{2.17}$$

where I is an identity matrix of $N \times N$ dimension, M is a diagonal matrix of $N \times N$ dimension containing values, denoted by m_{ii}, which are proportional to a conditional covariance $S_i|S_j$ and W is a weight matrix of $N \times N$ dimension with values, denoted by w_{ij}, representing the spatial correlation between region i and j and ρ is a representation of the amount of spatial dependence.

The joint multivariate Gaussian model is expressed in Spiegelhalter et al. (2003) as:

$$S_i|\mathbf{S}_{-i} \sim N\left(\mu_i + \sum_{j=1}^{N} \rho w_{ij}\big(S_j - \mu_j\big), \omega m_{ii}\right), \tag{2.18}$$

where $\mathbf{S}_{-i}$ denotes the whole vector $\mathbf{S}$ excluding the value of S_i. When there are no explanatory variables then $\mu_i = x_i'\beta = 0$ and Equation 2.18 can then be written as:

$$S_i|\mathbf{S}_{-i} \sim N\left(\sum_{j=1}^{N} \rho w_{ij} S_j, \omega m_{ii}\right). \tag{2.19}$$

Theorem 2.9. *When there are no explanatory variables then the joint distribution of the random effects are:*

$$\mathbf{S} \sim MVN(0, \omega\Sigma),$$

where $\omega\Sigma$ is given in Equation 2.17.

Proof. The Brook expansion developed by Brook (1964) may be written as:

$$\frac{\pi(\mathbf{z})}{\pi(\mathbf{x})} = \prod_{i=1}^{N} \frac{\pi\big(z_i|z_1, z_2, \ldots, z_{i-1}, x_{i+1}, \ldots, x_N\big)}{\pi\big(x_i|z_1, z_2, \ldots, z_{i-1}, x_{i+1}, \ldots, x_N\big)},$$

and thus:

$$\begin{aligned}
\frac{\pi(\mathbf{S})}{\pi(\mathbf{0})} &= \prod_{i=1}^{N} \frac{\pi\big(S_i|S_1,S_2,\ldots,S_{i-1},0_{i+1},\ldots,0_N\big)}{\pi\big(0_i|S_1,S_2,\ldots,S_{i-1},0_{i+1},\ldots,0_N\big)} \\
&= \prod_{i=1}^{N} \frac{exp\left\{-\frac{1}{2\omega m_{ii}}\left(S_i-\rho\sum_{j<i}w_{ij}S_j-\rho\sum_{j>i}0_j\right)^2\right\}}{exp\left\{-\frac{1}{2\omega m_{ii}}\left(0_i-\rho\sum_{j<i}w_{ij}S_j-\rho\sum_{j>i}0_j\right)^2\right\}} \\
&= \prod_{i=1}^{N} exp\left\{-\frac{1}{2\omega m_{ii}}\left[\left(S_i-\rho\sum_{j<i}w_{ij}S_j\right)^2-\left(\rho\sum_{j<i}w_{ij}S_j\right)^2\right]\right\} \\
&= \prod_{i=1}^{N} exp\left\{-\frac{1}{2\omega m_{ii}}\left[S_i^2-2S_i\rho\sum_{j<i}w_{ij}S_j+\left(\rho\sum_{j<i}w_{ij}S_j\right)^2-\left(\rho\sum_{j<i}w_{ij}S_j\right)^2\right]\right\} \\
&= \prod_{i=1}^{N} exp\left\{-\frac{1}{2\omega m_{ii}}\left[S_i^2-2\rho S_i\sum_{j<i}w_{ij}S_j\right]\right\} \\
&= exp\left\{-\frac{1}{2\omega\sum_{i=1}^{N}m_{ii}}\left[\sum_{i=1}^{N}S_i^2-2\rho\sum_{i=1}^{N}\sum_{j<i}S_iw_{ij}S_j\right]\right\}.
\end{aligned}$$

Since W is symmetric then $2\rho\sum_{i=1}^{N}\sum_{j<i}S_iw_{ij}S_j=\rho\sum_{i=1}^{N}\sum_{j=1}^{N}S_iw_{ij}S_j$ and thus:

$$\begin{aligned}
exp\left\{-\frac{1}{2\omega\sum_{i=1}^{N}m_{ii}}\left[\sum_{i=1}^{N}S_i^2-2\rho\sum_{i=1}^{N}\sum_{j<i}S_iw_{ij}S_j\right]\right\} &= exp\left\{-\frac{1}{2\omega\sum_{i=1}^{N}m_{ii}}\left[\sum_{i=1}^{N}S_i^2-\rho\sum_{i=1}^{N}\sum_{j=1}^{N}S_iw_{ij}S_j\right]\right\} \\
&= exp\left\{-\frac{1}{2\omega}\left[\sum_{i=1}^{N}S_i^2-\rho\sum_{i=1}^{N}\sum_{j=1}^{N}S_iw_{ij}S_j\right]\sum_{i=1}^{N}m_{ii}^{-1}\right\} \\
&= exp\left\{-\frac{1}{2}\left[S'\Big(I-\rho W\Big)\Big(\omega M\Big)^{-1}S\right]\right\} \\
&\sim MVN\Big(\mathbf{0},\omega\big(I-\rho W\big)^{-1}M\Big) \\
&\sim MVN\Big(\mathbf{0},\omega\Sigma\Big).
\end{aligned}$$

□

The use of the covariance matrix Σ requires the matrices W and M and the parameter ρ to be specified. The CAR model requires Σ to be symmetric positive definite such that the following conditions are met, as outlined in Spiegelhalter et al. (2003):

1. The matrix Σ must be symmetric such that $w_{ij}m_{jj}=w_{ji}m_{ii}$.

2. When $\rho = 0$ there is no spatial dependence.

3. The value of ρ must be in-between ρ_{min} and ρ_{max} such that ρ_{min}^{-1} and ρ_{max}^{-1} are lowest and highest eigenvalues of $M^{-\frac{1}{2}}WM^{\frac{1}{2}}$.

4. $Var\left(S_j|S_i\right) = \omega m_{ii} > 0$ such that $m_{ii} > 0$.

5. Since spatial dependence is mostly positive, ρ is constrained to be between 0 and ρ_{max}.

An alternative is to use the Simultaneous Autoregressive Model which does not require W to be symmetric.

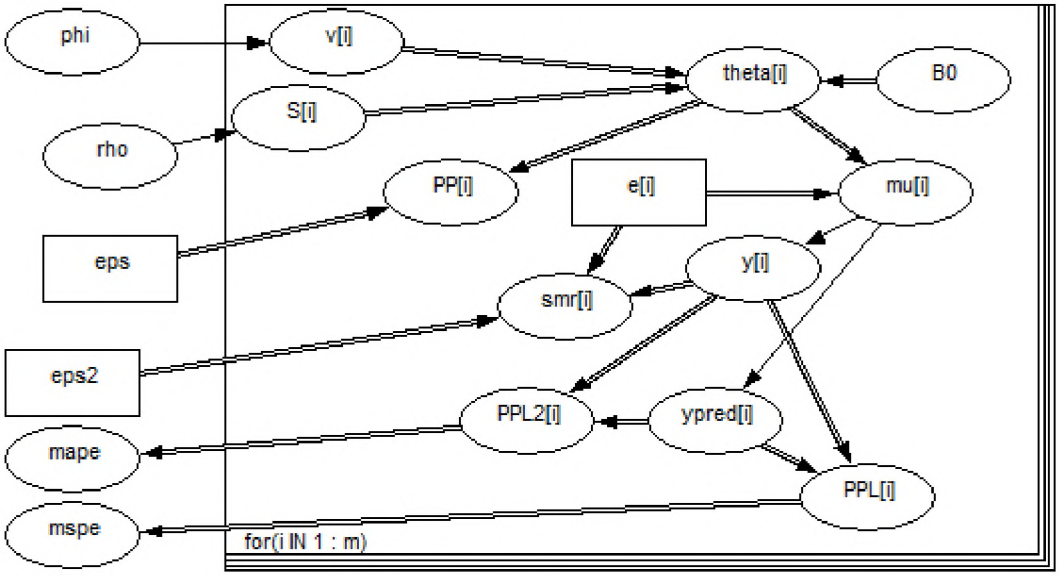


Figure 2.17: A simplified doodle for the model, not including all nodes.

2.12.3.1 Improper CAR (ICAR) Model

This model's foundation dates back to 1987 when Künsch (1987) developed lattice models for intrinsic autoregressions. These autoregressions were based on a two-dimensional lattice and could be applied to intrinsic models for which stationarity is only assumed when parameters change value. These models were built based on spatial distributions and permit the use of a "singular normal joint distribution" based on Lawson (2009).

Besag et al. (1991) show that the ICAR model is a CAR model where the covariance matrix Σ is semi-definite, whereas the CAR model has a positive definite covariance matrix as mentioned earlier. The ICAR model creates weights by:

$$w_{ij} = \begin{cases} \frac{1}{ne_i} & \text{if regions } i \text{ and } j \text{ are adjacent} \\ 0 & \text{Otherwise.} \end{cases}$$

and w_{ii} is also 0 and set $m_{ii} = \frac{1}{ne_i}$ where ne_i is the number of neighbours of region i. The ICAR model also requires:

$$\rho = 1 = \rho_{max}.$$

Since $\rho = 1,\ m_{ii} = \frac{1}{ne_i}$ and $w_{ij} = \frac{1}{ne_i}$ and since there are no explanatory variables in both data sets then Equation 2.19 will now change to the following equation for the ICAR model:

$$S_i|\mathbf{S}_{-i} \sim N\left(\sum_{j \epsilon ne_i}^{N} \frac{S_j}{ne_i}, \frac{\omega}{ne_i}\right).$$

WinBUGS® requires unnormalised weights and sets $Z_{ij} = 1$ for i and j adjacent regions and 0 otherwise and sets $w_{ij} = \frac{Z_{ij}}{Z_{ik}}$ where $Z_{ik} = \sum_{j=1}^{N} w_{ij}$.

A constraint which requires the random effects in the model sum to 0 is proposed by Besag & Kooperberg (1995) and includes an additional intercept term with a $U(-\infty, \infty)$ prior which is location invariant and gives the same result as an unconstrained parameterisation without an additional intercept term, this is modelled as β_0 in the Equation 2.20. The prior distribution for the parameter ω must be included in the model and be modelled in terms of precision as $\tau = \frac{1}{\omega}$. This prior is very sensitive because the posterior variance in the random effect will be affected by the prior variance as written by Spiegelhalter et al. (2003). The prior distribution is usually modelled by a gamma distribution whose parameters need to be selected sensibly as a large prior variance places most of the prior focus away from 0. Various suggestions are available for the parameters of the gamma distribution. The one proposed by Kelsall & Wakefield (1999) has parameters of $\alpha = 0.5$ and $\beta = 0.005$.

The model that is applied to both data sets is the CAR model which contains components from both the UH and CH models and is implemented in WinBUGS®. The model has the following parameters:

$$\begin{aligned} y_i &\backsim Pois(\mu_i) \text{ where } i = 1, 2, \ldots, N \text{ and } \mu_i = e_i\theta_i \\ L(\mu_i|y_i) &= \prod_{i=1}^{N} \frac{\mu_i^{y_i} exp(-\mu_i)}{y_i!} \\ E(y_i) &= e_i\theta_i = \mu_i \\ log(\mu_i) &= log(e_i) + \beta_0 + v_i + S_i, \end{aligned} \tag{2.20}$$

where θ_i is the relative risk in the i^{th} region, e_i is the expected number of events taking place in the i^{th} region, $v_i \sim norm(\mu = 0, \phi)$ is the uncorrelated heterogeneity, S_i is the ICAR model prior and β_0 is the intercept term due to the requirement that the random effects in the model must sum to 0. The model focuses on making inference on the relative risk of each region. The following link is assumed:

$$log(\mu_i) = \eta_i,$$

then

$$\mu_i = e_i^{\eta_i},$$

which results in the Bayesian linear model given by:

$$\eta_i = x_i'\beta + \beta_0 + v_i + S_i.$$

The relative risk is modelled by:

$$\theta_i = exp\Big[\beta_0 + v_i + S_i\Big],$$

where $x_i'\beta$ is not modelled as there are no explanatory variables in the data.

The prior distribution for the vector $\boldsymbol{S}$ is written by Lawson (2009) as:

$$\pi(\mathbf{S}|\omega) \propto \frac{1}{\omega^{\frac{N}{2}}} exp\left\{-\frac{1}{2\omega}\sum_i\sum_{j \epsilon ne_i}\Big(S_i - S_j\Big)^2\right\},$$

which is an application of a "Markov random field." There are various methods for weighting schemes for regions.

The zero-mean Gaussian prior distribution was assigned by Besag et al. (1991) to the uncorrelated heterogeneity v_i in each region. This prior is written as:

$$\pi(\mathbf{v}) \propto \phi^{-\frac{N}{2}} exp\left\{-\frac{1}{2\phi}\sum_{i=1}^{N} v_i^2\right\},$$

where ω and ϕ have these priors for the South African models:

$$\omega \sim gamma\Big(\alpha_1, \beta_1\Big)$$

$$\phi \sim gamma\Big(\alpha_2, \beta_2\Big),$$

and the following priors for the Georgia models:

Model 1:

$$\begin{aligned} \omega &\sim gamma\Big(\alpha_3, \beta_3\Big) \\ \phi &\sim gamma(\alpha_4, \beta_4). \end{aligned}$$

Model 2:

$$\begin{aligned}\omega &\sim gamma\Big(\alpha_5, \beta_5\Big)\\ \phi &\sim gamma(\alpha_6, \beta_6).\end{aligned}$$

Model 3:

$$\begin{aligned}\omega &\sim gamma\Big(\alpha_7, \beta_7\Big)\\ \phi &= constant.\end{aligned}$$

The posterior distribution for the South African models have the form:

$$\begin{aligned}\pi\Big(\boldsymbol{S}, \boldsymbol{v}, \omega, \phi | y_i\Big) \propto & \prod_{i=1}^{N} \left\{ \frac{exp\Big(-e_i\theta_i\Big)\Big(e_i\theta_i\Big)^{y_i}}{y_i!} \right\} \times \frac{1}{\omega^{\frac{N}{2}}} exp\left\{ -\frac{1}{2\omega} \sum_i \sum_{j \epsilon ne_i} \Big(S_i - S_j\Big)^2 \right\} \\ & \times \phi^{-\frac{N}{2}} exp\left\{ -\frac{1}{2\phi} \sum_{i=1}^{N} v_i^2 \right\} \times gamma\Big(\alpha_1, \beta_1\Big) \times gamma\Big(\alpha_2, \beta_2\Big).\end{aligned}$$

The posterior distribution for Georgia model 1 is as follows:

$$\begin{aligned}\pi\Big(\boldsymbol{S}, \boldsymbol{v}, \omega, \phi | y_i\Big) \propto & \prod_{i=1}^{N} \left\{ \frac{exp\Big(-e_i\theta_i\Big)\Big(e_i\theta_i\Big)^{y_i}}{y_i!} \right\} \times \frac{1}{\omega^{\frac{N}{2}}} exp\left\{ -\frac{1}{2\omega} \sum_i \sum_{j \epsilon ne_i} \Big(S_i - S_j\Big)^2 \right\} \\ & \times \phi^{-\frac{N}{2}} exp\left\{ -\frac{1}{2\phi} \sum_{i=1}^{N} v_i^2 \right\} \times gamma\Big(\alpha_3, \beta_3\Big) \times gamma(\alpha_4, \beta_4).\end{aligned}$$

Model 2:

$$\begin{aligned}\pi\Big(\boldsymbol{S}, \boldsymbol{v}, \omega, \phi | y_i\Big) \propto & \prod_{i=1}^{N} \left\{ \frac{exp\Big(-e_i\theta_i\Big)\Big(e_i\theta_i\Big)^{y_i}}{y_i!} \right\} \times \frac{1}{\omega^{\frac{N}{2}}} exp\left\{ -\frac{1}{2\omega} \sum_i \sum_{j \epsilon ne_i} \Big(S_i - S_j\Big)^2 \right\} \\ & \times \phi^{-\frac{N}{2}} exp\left\{ -\frac{1}{2\phi} \sum_{i=1}^{N} v_i^2 \right\} \times gamma\Big(\alpha_5, \beta_5\Big) \times gamma(\alpha_6, \beta_6).\end{aligned}$$

Model 3:

$$\pi\Big(\boldsymbol{S},\boldsymbol{v},\omega,\phi|y_i\Big) \propto \prod_{i=1}^{N}\left\{\frac{exp\Big(-e_i\theta_i\Big)\Big(e_i\theta_i\Big)^{y_i}}{y_i!}\right\} \times \frac{1}{\omega^{\frac{N}{2}}}exp\left\{-\frac{1}{2\omega}\sum_i\sum_{j\epsilon ne_i}\Big(S_i-S_j\Big)^2\right\}$$
$$\times\phi^{-\frac{N}{2}}exp\left\{-\frac{1}{2\phi}\sum_{i=1}^{N}v_i^2\right\}\times gamma\Big(\alpha_7,\beta_7\Big)\times constant.$$

The parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, are determined in Section 2.9.4. The advantage of using such a model is shown in Lawson (2009) as conditional moments:

$$\begin{aligned} E\Big(S_i|\mathbf{S}_{-i}\Big) &= \bar{S}_i \\ var\Big(S_i|\mathbf{S}_{-i}\Big) &= \frac{\omega}{ne_i} \\ \Big[S_i|\mathbf{S}_{-i}\Big] &\sim N\Big(\bar{S}_i,\frac{\omega}{ne_i}\Big), \end{aligned}$$

where $\bar{S}_i = \sum_{j\epsilon ne_i}\frac{S_j}{ne_i}$, the average calculated over the districts of the i^{th} region.

Thus an advantage of using this model is that the conditional moments are easy to calculate and are functions of the regions in the study.

The justification for using both CH and UH components in the model is provided in Lawson (2009). The CH and UH components must be used in the model because unobserved effects within the region under study can arise in different forms. The UH effect is included in the model to model uncorrelated additional variation. In the absence of prior information, there is little justification not to include both effects, and it is straightforward to include both effects in the form of an additive random effect in the model. These effects may not be identified, but focus may be on the "total effect of the unobserved confounding," then the sum of these effects can be identified and forms part of the model.

The intraclass correlation coefficient (ICC) may be calculated from the variance components of the UH and CH components as:

Let ϕ_ν and ϕ_S denote the variance of the UH and CH components respectively, then the "relative variance contribution/intraclass correlation" is given by $\frac{\phi_\nu}{\phi_\nu+\phi_S}$, as in by Lawson (2009). This "relative variance contribution" is only useful when the components may be identified.

The WinBUGS® function for the ICAR model is the *car.normal* function. The following parameters are required for the ICAR model as is copied from Spiegelhalter et al. (2003):

$$\boldsymbol{S}[1:N] \sim car.normal(\boldsymbol{adj}[],\boldsymbol{weights}[],\boldsymbol{num}[],omega),$$

where:

$\boldsymbol{adj}[]$ is a vector containing the identification numbers of the neighbours of each region and is

generated using the adjacency matrix in the mapping tool menu in GeoBUGS®(2003).

$\boldsymbol{weights}[]$ is a vector containing the unnormalised weights for each pair of regions and having the same length as the $\boldsymbol{adj}[]$ vector. This is generated as a vector of 1's by assigning $m_{ij} = 1$ $(w_{ij} = \frac{1}{ne_i})$ if regions i and j are adjacent and 0 otherwise. This is the standard CAR model proposed by Besag et al. (1991).

$\boldsymbol{num}[]$ is a vector the same length as the number of regions in the study and contains the number of neighbours (ne_i) for each region. This vector is created using the adjacency matrix in the mapping tool menu in GeoBUGS®.

$omega$ is a value which is the precision of the CAR prior or the inverse scale parameter when using the Laplace prior. This may be assigned a gamma prior distribution in the model.

The intercept term β_0 is assigned a flat prior distribution and must be used in a CAR model that has random effects.

2.12.3.2 Proper CAR (PCAR) Model

Another form of the Gaussian Markov random field (GMRF) model is the PCAR model. Suppose ρ is as defined earlier and to ensure definiteness of Σ, the covariance matrix, ρ must lie in the interval ρ_{min} and ρ_{max} such that ρ_{min}^{-1} and ρ_{max}^{-1} are the lowest and highest eigenvalues of $M^{-\frac{1}{2}}WM^{\frac{1}{2}}$, as mentioned earlier. Thus the range of ρ is a function of the eigenvalues of the matrix $M^{-\frac{1}{2}}WM^{\frac{1}{2}}$. This is different to the ICAR model where ρ is fixed, as explained in Spiegelhalter et al. (2003). The distribution which arises is thus proper assuming that the properties of M and W are met. A straight forward uniform distribution, $U\big(\rho_{min}, \rho_{max}\big)$, may be given as a hyperprior for ρ, as mentioned by Besag & Kooperberg (1995). The biggest advantage of using this type of model is that the results will be similar to those produced by the fully specified Gaussian covariance model because the model does not have to perform a matrix inversion during sampling. The model may also be used as a data likelihood as it is proper and the variance and correlation parameters are specified, and this is in contrast to the ICAR model which is improper and therefore cannot be used as a data likelihood. There are two types of specifications which will be considered in this research. The specifications are for the elements of the matrices M and W and the parameter ρ. The first specification is made by Besag et al. (1991) and is used in the ICAR model, this specification is:

$$w_{ij} = \begin{cases} \frac{1}{ne_i} & \text{if regions } i \text{ and } j \text{ are adjacent.} \\ 0 & \text{Otherwise.} \end{cases}$$

and

$$m_{ii} = \frac{1}{ne_i},$$

$$\rho = 1 = \rho_{max}.$$

The second specification, which is used in the PCAR model, is proposed by Cressie & Chan (1989) and Stern & Cressie (1999):

$$w_{ij} = \begin{cases} \left(\frac{e_i}{e_j}\right)^{\frac{1}{2}} & \text{if regions } i \text{ and } j \text{ are adjacent} \\ 0 & \text{Otherwise.} \end{cases} \tag{2.21}$$

and

$$m_{ii} = \frac{1}{e_i}, \tag{2.22}$$

$$\rho \in \left(\rho_{min}, \rho_{max}\right),$$

where e_i is the expected number of events at the i^{th} region.

The model with no trend is then written as:

$$S_i|\mathbf{S}_{-i} \sim N\left(\rho \sum_{j \epsilon n e_i}^{N} \left(\frac{e_i}{e_j}\right)^{\frac{1}{2}} S_j, \frac{\omega}{e_i}\right).$$

The WinBUGS® function for the PCAR model is the *car.proper* function. The following parameters are required for the PCAR model, as is copied from Spiegelhalter et al. (2003):

$$\boldsymbol{S}[1:N] \sim car.proper(\boldsymbol{mu}[], \boldsymbol{W}[], \boldsymbol{adj}[], \boldsymbol{num}[], \boldsymbol{M}[], omega, \rho),$$

where:

$\boldsymbol{mu}[]$ is a vector containing the means for each region, this may be given a prior distribution, defined in the data or calculated within the model.

$\boldsymbol{adj}[]$, $omega$ and $\boldsymbol{num}[]$ are the same as in the ICAR model.

$\boldsymbol{W}[]$ is a vector of normalised weights of each pair of regions. The ICAR model used unnormalised weights.

$\boldsymbol{M}[]$ is a vector containing the diagonal values of M denoted by m_{ii} and has length N.

ρ is a value which specifies the strength of the spatial dependence and is constrained by the highest and lowest eigenvalues of $M^{-\frac{1}{2}} W M^{\frac{1}{2}}$.

The bounds are created by using the parameters of the PCAR model as:

$$min(\boldsymbol{W}[], \boldsymbol{adj}[], \boldsymbol{num}[], \boldsymbol{M}[]) \text{ and } max(\boldsymbol{W}[], \boldsymbol{adj}[], \boldsymbol{num}[], \boldsymbol{M}[]).$$

2.12.3.3 Differences Between the Choice of Covariance matrix in ICAR and PCAR models

The difference between the choice of $w_{ij} = \frac{1}{ne_i}$ and $w_{ij} = \left(\frac{e_i}{e_j}\right)^{\frac{1}{2}}$ is that $\sum_{j=1}^{n} w_{ij} = 1$ under $w_{ij} = \frac{1}{ne_i}$ whereas in $w_{ij} = \left(\frac{e_i}{e_j}\right)^{\frac{1}{2}}$ this is not the case. The difference between choosing $\rho = \rho_{max}$ in the ICAR model gives the highest spatial relationship but the covariance matrix is singular at this value, resulting in an improper prior distribution for the relative risk. Letting $\rho \in (\rho_{min}, \rho_{max})$ change according to the data results in a positive-definite matrix. The conditional correlation is also different as shown by Stern & Cressie (1999) with differences resulting from the choice of ρ and w_{ij}. The partial or conditional correlation is written as: $\rho^2 w_{ij} w_{ji} = corr^2\{\theta_i \theta_j | \theta_{\mathbf{i-j}}\}$. When $w_{ij} = \frac{1}{ne_i}$ and $\rho = 1$ in the ICAR model then: $\rho^2 w_{ij} w_{ji} = \left(\frac{1}{ne_i}\right)\left(\frac{1}{ne_j}\right)$ such that the partial correlation is a function of the number of neighbours and gives results similar to those given by the product of the partial variances. When $w_{ij} = \left(\frac{e_i}{e_j}\right)^{\frac{1}{2}}$ and $\rho \in (\rho_{mix}, \rho_{max})$ then the partial correlation is: $\rho^2 w_{ij} w_{ji} = \rho^2$ for when $\rho \in (\rho_{min}, \rho_{max})$ is independent of the form of m_{ii}.

2.12.4 Poisson Process Model

The Poisson process model is used by Bernardinelli et al. (1995) when a disease is not contagious and does not often occur in the regions under study. Let $\{D_i\}$, $i = 1, 2, \ldots N$, denote the set of all events taking place in all regions under the study and let e_i denote the expected number of events in region i. Events recorded in the study should be addressed locations of each event, as stated by Lawson (2009). It is assumed that events are Geo-coded as a point with reference to the same scale as the entire region under the study and it is also assumed that all events within a region are documented.

Lawson (2009) does not consider incomplete documentation of events and model's data documented as a heterogeneous Poisson process with a first order intensity denoted by $\lambda(s)$. Another assumption which needs to be made to use the model is that events are assumed to be "independently spatially-distributed" and ruled by the first order intensity, $\lambda(s)$. The unconditional likelihood, which requires events to be independent, is given in Lawson (2009) as:

$$L\Big(\{D_i\}|\boldsymbol{\psi}\Big) = \frac{1}{m!}\prod_{i=1}^{N}\lambda(D_i|\boldsymbol{\psi})exp\Big(-\wedge_T\Big),$$

where

$$\wedge_T \approx \sum_{i=1}^{N} w_i \lambda\big(D_i|\boldsymbol{\psi}\big),$$

where T is the region under study and $\wedge_T$ is the integral of the intensity in the region, $\lambda\big(D_i|\boldsymbol{\psi}\big)$ is the first order intensity of events in the i^{th} region and $\boldsymbol{\psi}$ is a vector of parameters. When likelihood based inference is required in the study, the likelihood is maximised with reference to the parameters in the vector $\boldsymbol{\psi}$.

This method may prove to be challenging as the integral $\wedge_T$ may not always be easily solved. Lawson (2009) proposes the method of numerical integration to solve this problem. Numerical integration is used along with assigning a weighting method to the likelihood. An example of a weighting method used by Lawson (2009) is:

$$L\bigg(\{D_i\}\,|\boldsymbol{\psi}\bigg) \;=\; \sum_{i=1}^{N} ln\bigg(\lambda\big(D_i|\boldsymbol{\psi}\big)\bigg) - \wedge_T,$$

where w_i is the weight and $\wedge_T \approx \sum_{i=1}^{N} w_i\lambda\big(D_i|\boldsymbol{\psi}\big)$. $\wedge_T$ is an approximation and hence a more accurate method needs to be used. Mark & Turner (1992) approximate the integral in this equation:

$$l\big(\theta; x_1, x_2, \ldots, x_n; T\big) \;=\; \sum_{i=1}^{N} log\bigg(\lambda_\theta\Big(x_i|F_{x_i}\Big)\bigg) - \int_0^T \lambda_\theta\Big(x|F_x\Big)dx,$$

by using a weighted sum:

$$\int_0^T \lambda_\theta\Big(x|F_x\Big)dx \;=\; \sum_{j=1}^{N} w_j\lambda_\theta\Big(D_j|F_{D_j}\Big).$$

Substitution then results in the following equation:

$$l(\theta) \;=\; \sum_{j=1}^{N} w_j\left(\frac{N_j}{w_j} ln\lambda_\theta\Big(D_j|F_{D_j}\Big) - \lambda_\theta\Big(D_j|F_{D_j}\Big)\right).$$

It can be shown that this equation represents the "weighted log likelihood of independent Poisson variates with parameters $\lambda_\theta\Big(D_j|F_{D_j}\Big)$."

Lawson (2009) applies this method to spatial data as:

$$L\bigg(\{D_i\}\,|\boldsymbol{\psi}\bigg) \;=\; \sum_{k=1}^{N} w_j\left(\frac{I_k}{w_k} ln\lambda\Big(D_k|\boldsymbol{\psi}\Big) - \lambda\Big(D_k|\boldsymbol{\psi}\Big)\right),$$

where

$$\int_T \lambda(u|\boldsymbol{\psi})du \;=\; \sum_{k=1}^{N} w_k \lambda\Big(D_k|\boldsymbol{\psi}\Big),$$

where I_k is an indicator function and $I_k = 1$ for an event and 0 otherwise. The intensity $\lambda(S|\boldsymbol{\psi})$ may also be modelled by specifying the intensity in terms of two functions. The two functions are important as one function models the risk associated with the study region and the other function accounts for the core population for the study region.

2.12.5 Conditional Logistic Model

Conditional logistic regression is said by Greenland et al. (2000) to have been produced to reduce "sparse-data biases" resulting from the use of ordinary logistic regression. The conditional logistic regression is used in large samples of data but may show a significant amount of bias when there are too many parameters in the model or when there is no consistency in the matched sets. Lawson (2009) says that conditional logistic regression is used when bivariate cases arise in data, and these cases require conditional inference.

In this model, there are events which fall under control, denoted by $y_i = 0$, and case events denoted by $y_i = 1$, where y_i is a binary variable linked with each location or address. The case events are defined by Lawson (2009) as U_i where $i = 1, 2, \ldots, b$ and control events are U_i where $i = b+1, \ldots, N$ where $N = n+b$, the total number of events in the study. The events (case and control) are assumed to be distributed as a heterogeneous Poisson process model with an associated intensity parameter given as $\lambda(U|\boldsymbol{\psi})$ for case events and $\lambda_0(U|\boldsymbol{\psi_0})$ for control events. The addition of these events results in a heterogeneous Poisson process with an associated intensity parameter given by Lawson (2009) as $\lambda_0\Big(U|\boldsymbol{\psi_0}\Big) + \lambda(U|\boldsymbol{\psi}) = \lambda_0\Big(U|\boldsymbol{\psi_0}\Big)\Big(1 + \lambda_1(U|\boldsymbol{\psi_1})\Big)$. The case events and control events conditional probability at location i is given by Lawson (2009) as:

$$\begin{aligned} P\big(y_i = 1\big) &= \frac{\lambda_0\big(U_i|\boldsymbol{\psi_0}\big)\lambda_1\big(U_i|\boldsymbol{\psi_1}\big)}{\lambda_0\Big(U_i|\boldsymbol{\psi_0}\Big)\Big(1 + \lambda_1\big(U_i|\boldsymbol{\psi_1}\big)\Big)} \\ &= \frac{\lambda_1\big(U_i|\boldsymbol{\psi_1}\big)}{1 + \lambda_1\big(U_i|\boldsymbol{\psi_1}\big)} \\ &= p_i \end{aligned} \tag{2.23}$$

and,

$$\begin{aligned} P(y_i = 0) &= \frac{1}{1 + \lambda_1(U_i|\boldsymbol{\psi_1})} \\ &= 1 - p_i. \end{aligned}$$

If Equation 2.23 is used in a model then the likelihood function of case and control events becomes:

$$\begin{aligned} L(\boldsymbol{\psi_1}|U) &= \prod_{i\,\epsilon\,cases} p_i \prod_{i\,\epsilon\,controls} 1 - p_i \\ &= \prod_{i=1}^{N} \left[\frac{exp(\eta_i)^{y_i}}{1 + exp(\eta_i)} \right], \end{aligned}$$

where $\eta_i = x_i^{'}\beta$ and $x_i^{'}$ is the i^{th} row of a design matrix containing covariates and β is the associated vector of parameters with length p.

2.12.6 Binomial Model

The binomial model is used when a study is based on small regions and count data of events is measured within each region. Let the study be made up of a total of m small regions with the observed count of each region denoted by y_i. The assumption of a finite population in each region is made by Lawson (2009) with the population of the m regions denoted by n_i where $i = 1, 2, \ldots, m$. The model is then the observed counts conditional on the population count. Then the model is given in Lawson (2009) as:

$$y_i \sim bin(p_i,\ n_i),$$

where p_i is the probability of an event taking place in region i, $i = 1, 2, \ldots, m$, and with a likelihood given as:

$$L(y_i|p_i, n_i) = \prod_{i=1}^{m} \begin{pmatrix} n_i \\ p_i \end{pmatrix} p_i^{y_i}(1 - p_i)^{n_i - y_i}.$$

A link function for the model is usually chosen. The more popular link function is the logit link function given in Lawson (2009) as:

$$p_i = \frac{exp(\eta_i)}{1 + exp(\eta_i)}.$$

2.13 Model Diagnostics

There are many different types of model diagnostics which are applied in different situations and dependent on the problem at hand. A misspecified model results in poor estimation and usually has no

or limited use.

2.13.1 Deviance Information Criterion (DIC)

The Deviance Information Criterion (DIC) was proposed by Spiegelhalter et al. (2002) and is commonly used in goodness-of-fit (GOF) measures. The DIC is directly available in WinBUGS® and is calculated by:

$$DIC = 2E_{\theta|y}(D) - D\Big\{E_{\theta|y}(\theta)\Big\},$$

where D is the deviance and is given by Lawson (2009) as:

$$D = -2\left[l\Big(y|\hat{\theta}_{fit}\Big) - l\Big(y|\hat{\theta}_{sat}\Big)\right],$$

and compares a fitted model to a saturated model and y is the data in the model. The DIC may also be calculated by:

$$DIC = \bar{D} + p\hat{D} = 2\bar{D} - \hat{D}\Big(\hat{\theta}\Big),$$

where $p\hat{D}$ is the effective number of parameters and $\bar{D}$ is the average deviance calculated by Lawson (2009) as:

$$\bar{D} = -2\sum_{g=1}^{G}\Big(l\big(y|\theta^g\big)/G\Big),$$

where G is the number of posterior samples and $\hat{D}\Big(\hat{\theta}\Big)$ is the deviance of the posterior expected parameter estimate $\hat{\theta}$, and is calculated by:

$$\hat{D}\Big(\hat{\theta}\Big) = -2l\Big(y|\hat{\theta}\Big).$$

Lawson (2009) states that models may have an incorrect pD when the overdispersion in the model results in $\hat{D}\Big(\hat{\theta}\Big) > \bar{D}$. This may also be caused by the poor choice in the hyperpriors of the hierarchical model and poor choices of variances in prior distributions. A more robust method to calculate the effective number of parameters based on the posterior variance of the deviance is given in Gelman et al. (2014) as:

$$\tilde{p}D = \frac{1}{2}\frac{1}{G-1}\sum_{g=1}^{G}\Big(D\big(y,\theta^g\big) - \bar{D}\Big)^2,$$

which is derived from the output of an MCMC chain. Another method which can be calculated from the output of the MCMC chain is an estimator of the variance and is given by Lawson (2009) as:

$$\hat{var}(D) = \frac{1}{G-1}\sum_{g=1}^{G}\Big(\hat{D}\big(\theta^g\big) - \bar{D}\Big)^2 = 2\tilde{p}D,$$

and then the DIC becomes:

$$DIC = \bar{D} + \hat{var}(D).$$

2.13.2 Posterior Predictive Loss

Gelfand & Ghosh (1998) introduced a criterion for model choice by minimising the posterior loss for a known model and then selecting the model under the study which minimises the criterion. Let y_i^* denote the i^{th} observation from the predictive data and let $\theta^{(j)}$ denote all the parameters at the j^{th} iteration in the MCMC. Then y^* is calculated by Lawson (2009) as:

$$\pi\big(y_i^*|\mathbf{y}\big) = \int \pi\big(y_i^*|\theta^{(j)}\big)\pi\big(\theta^{(j)}|y\big)d\theta^{(j)}.$$

This is calculated for the model with a Poisson likelihood by assigning: $y_i^* \sim Pois\big(e_i\theta_i^{(j)}\big)$. The squared error loss can then be defined as:

$$L_*\big(y, y^*\big) = \big(y - y^*\big)^2,$$

where $L_*\big(y, y^*\big) = f\big(y, y^*\big)$.

The mean squared predictive error (MSPE) can then be written by Lawson (2009) as:

$$MSPE_j = \sum_j \frac{\big(y_i - y_{ij}^*\big)^2}{n_T},$$

where n_T is the sample size. The overall MSPE is then calculated by:

$$MSPE = \sum_i \sum_j \frac{\big(y_i - y_{ij}^*\big)^2}{n_T \times G},$$

where G is the "sampler sample size." The mean absolute predictive error (MAPE) can also be calculated as:

$$MAPE_j = \sum_i \frac{|y_i - y_{ij}^*|}{n_T}$$

and the overall MAPE is then:

$$MAPE = \sum_i \sum_j \frac{|y - y_{ij}^*|}{n_T \times G}.$$

The model with a smaller MAPE or MSPE would be the better model to use. The MAPE and MSPE are calculated for all the models in this study and are based on the method used by Lawson (2009).

2.13.3 Assessing Relative Risk

The risk of each region may be assessed using an exceedance probability. The exceedance probability may also be used to detect and assess unexpected clustering or accumulation of the disease in the model. The most basic form of the exceedance probability is denoted by Lawson (2009) as:

$$q_i^c = P\big(\theta_i > c\big).$$

This is thus an estimate of the number of times the relative risk surpasses the relative risk of 1, $\theta_i = 1$, which is referred to as the "null risk value." The exceedance probability can be a tool used to check for 'hot- spot' clusters and regions which have a higher risk than usual. After a MCMC has run and convergence for the sample has been reached, let the converged sample be denoted by: $\left\{\theta^{N_B+1}, \theta^{N_B+2}, \ldots, \theta^{N_B+n_p}\right\}$, as in Lawson (2009). Then the exceedance probabilities are calculated as:

$$\hat{q_i^c} = \sum_{g=N_B+1}^{N_B+n_p} \frac{I\big(\theta_i^{(g)} > c\big)}{G},$$

where $G = n_p$. The average of $\hat{q_i^c}$ for each region will yield the posterior expected exceedance probability which is denoted as PP. High values of $\hat{q_i^c}$ indicate that a region has a high and unusual risk. Lawson (2009) warns that the PP may be model dependent which is the result of the PP being applied to any underlying model.

The value of the parameter c in $\hat{q_i^c}$ must be specified, which may cause a problem in the calculation of $\hat{q_i^c}$. Let the exceedance probability of interest be denoted by $\hat{P}$, this is the probability in which is said to be unusual in the model. Then there is a trade-off between c and $\hat{P}$. Different values of c will lead to different unusual exceedance probabilities which in turn affect $\hat{P}$ as exceedance probabilities are compared to $\hat{P}$ to determine if the exceedance probability is unusual. Thus the value of $\hat{P}$ or c must be fixed.

2.13.4 Residuals

2.13.4.1 Overview

A common technique used in a model goodness-of-fit is the analysis of the residuals of the model. The residuals represent the difference between the observed value and the expected value, this difference is known as the deviance, as defined by Lunn et al. (2012). Residuals may be used to analyse various short falls in the model, some of which are named in Lunn et al. (2012) as: autocorrelation, the shape of the distribution and the fit of the model. The general case of a residual is defined as:

$$r_{Ri} = y_i - \hat{y}_i, \tag{2.24}$$

where $\hat{y}_i$ is the fitted value of the model.

2.13.4.2 Pearson Residuals

The standardised form of the Pearson residual is well-known and defined as:

$$r_{pi} = \frac{y_i - E\Big(y_i|\theta\Big)}{\sqrt{var\Big(y_i|\theta\Big)}},$$

where y_i is the i^{th} observed value, $E\big(y_i|\theta\big)$ is the expected value of the i^{th} observed value given the parameter θ and $var\Big(y_i|\theta\Big)$ is the variance of the i^{th} observed value given the parameter θ. The Pearson residual, r_p, has a mean of 0 and a variance of 1 and thus the residuals are assumed to occur between -2 and 2, as explained by Lunn et al. (2012).

2.13.4.3 Bayesian Residuals

The Bayesian residual is defined by Lawson (2009) as:

$$r_{Bi} = y_i - \frac{1}{G}\sum_{g=1}^{G} E\Big(y_i|\boldsymbol{\theta}_i^{(g)}\Big),$$

where $\boldsymbol{\theta}_i^{(g)}$ is a vector of values resulting from the sampling of the posterior distribution and $E\Big(y_i|\boldsymbol{\theta}_i^{(g)}\Big)$ is the expected value resulting from the output of the posterior distribution.

In the case of a model with a Poisson likelihood with mean parameter $e_i\theta_i$, the Bayesian residual can be approximated by:

$$r_{Bi} = y_i - \frac{1}{G}\sum_{g=1}^{G} e_i\theta_i$$

for the case when a constant region rate is applied to the model. Another parameterisation is to use the posterior expected value, $e_i\hat{\theta}_i$, as the fitted value and then the Bayesian residual is calculated in Lawson (2009) as:

$$r_{Bi} = y_i - e_i\hat{\theta}_i,$$

which follows from Equation 2.24.

2.13.4.4 Deviance Residuals

The deviance residual is defined by Lunn et al. (2012) as:

$$r_{di} = sign_i\sqrt{D_i}, \tag{2.25}$$

where $sign_i$ is the i^{th} sign of the result of $y_i - \hat{y}_i$ and D_i is the i^{th} standardised deviance defined as:

$$D(\boldsymbol{\theta}) = \sum D_i(\boldsymbol{\theta}) = -2log\Big[\pi\big(y|\boldsymbol{\theta}\big)\Big] + 2log\Big[\pi\big(y|\hat{\theta}_i(y)\big)\Big], \tag{2.26}$$

where $\hat{\theta}_i(y)$ is the "saturated estimate" as in Lunn et al. (2012).

Theorem 2.10. *The standardised deviance for the Poisson distribution when $y_i \sim Poisson(\theta_i)$ is:*

$$D(\boldsymbol{\theta}) = 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\hat{\theta}_i}\right) - \Big(y_i - \hat{\theta}_i\Big)\right]. \tag{2.27}$$

Proof. Let $\hat{\theta}_i$ denote the i^{th} maximum likelihood estimate of the fitted model and let $\tilde{\theta}_i = y_i$ under the saturated model. The likelihood of the fitted model is then:

$$L\Big(\hat{\theta}_i|y_i\Big) = \prod_{i=1}^{N}\frac{exp\Big(-\hat{\theta}_i\Big)\hat{\theta}_i^{y_i}}{y_i!}$$

and the likelihood of the saturated model is:

$$L\Big(\tilde{\theta}_i|y_i\Big) = \prod_{i=1}^{N}\frac{exp\Big(-\tilde{\theta}_i\Big)\tilde{\theta}_i{}^{y_i}}{y_i!}.$$

The log likelihood of the fitted model is then:

$$l\Big(\hat{\theta}_i|y_i\Big) = \sum_{i=1}^{N}\Big[y_i log\Big(\hat{\theta}_i\Big) - \hat{\theta}_i - log\Big(y_i!\Big)\Big]$$

and the log likelihood of the saturated model is:

$$l\Big(\tilde{\theta}_i|y_i\Big) = \sum_{i=1}^{N}\Big[y_i log\Big(\tilde{\theta}_i\Big) - \tilde{\theta}_i - log\Big(y_i!\Big)\Big].$$

Substituting the log likelihoods into Equation 2.26 results in:

$$\begin{aligned}
D(\boldsymbol{\theta}) &= -2\sum_{i=1}^{n}\Big[y_i log\Big(\hat{\theta}_i\Big) - \hat{\theta}_i - log\Big(y_i!\Big)\Big] + 2\sum_{i=1}^{n}\Big[y_i log\Big(\tilde{\theta}_i\Big) - \tilde{\theta}_i - log\Big(y_i!\Big)\Big] \\
&= 2\sum_{i=1}^{n}\Big[log\Big(y_i!\Big) - y_i log\Big(\hat{\theta}_i\Big) + \hat{\theta}_i + y_i log\Big(\tilde{\theta}_i\Big) - \tilde{\theta}_i - log\Big(y_i!\Big)\Big] \\
&= 2\sum_{i=1}^{n}\left[\hat{\theta}_i + y_i log\left(\frac{\tilde{\theta}_i}{\hat{\theta}_i}\right) - \tilde{\theta}_i\right].
\end{aligned}$$

Since $\tilde{\theta}_i = y_i$ under the saturated model, then the deviance is:

$$\begin{aligned} D(\boldsymbol{\theta}) &= 2\sum_{i=1}^{n}\left[\hat{\theta}_i + y_i log\left(\frac{y_i}{\hat{\theta}_i}\right) - y_i\right] \\ &= 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\hat{\theta}_i}\right) - \left(y_i - \hat{\theta}_i\right)\right]. \end{aligned}$$

□

Chapter 3

Data Analysis

3.1 Data Protection and Privacy

The sensitive nature of the data used in health research is critical. Health data may contain significant amounts of personal information which require confidentiality. Medical records are often difficult to access due to current legislation. Government and Departments of Health may also be reluctant to release this data as it has a susceptible nature. Elliott & Wartenberg (2004) explain that current legislation in the United States is the Privacy Rule introduced by Department of Health and Human Services (2002) which came into effect from 2003 and is part of the Health Insurance Portability and Accountability Act (1996). This legislation prevents epidemiology studies somewhat as data cannot be used if the initial intention that it was collected for was for a different purpose, although consent may be given to use the data depending on the circumstances. The legislation in the United Kingdom allows for this data to be used although security and safe practices to protect the data must be put into place. The nature of epidemiology studies makes them an important part of health research and as such, legislations are often changed.

Another problem in spatial studies, according to Elliott & Wartenberg (2004), is that certain studies break the privacy of the communities and neighbourhoods in their study regions. For example, regions with a high disease rate may cause many residents to move to other communities or regions or the property values of that region will decrease significantly. Thus the introduction of legislation into these studies is vital for the protection of data but also for the growth of medical research.

Historically, legislation for the protection of personal information was poor in South Africa. The growth of industries, information technology and storage options resulted in a change in South African Law to protect private information. The introduction of the Protection of Personal Information Act No. 4 introduced in 2013, according to South African Government (2013), changed the way in which personal data is protected. The Act states that personal information relates to any information regarding "education, medical, financial, criminal or employment history of a person" and includes many other topics. Other legislation relating to the confidentiality of medical records are given by

DLA Cliffe Dekker Hofmeyr (2012) under the Health Professions Act, No. 56 (HPA) introduced in 1974. This Act relates to health care providers and provides guidelines and rules which are set by the Health Professions Council of South Africa (HPCSA). The Act prevents medical staff from sharing patient information without the patient's consent. DLA Cliffe Dekker Hofmeyr (2012) says that the information can be shared when there are:

1. "Statutory provisions" in place.
2. Instructions to do so by a court or law.
3. Justifications that the information is in the public interest.

HPCSA also provides guidelines for the safe storage of patient information as is required by the National Health Act, No. 61 introduced in 2003. Other Acts and legislation relating to privacy and data protection can be found in the South African Law Reform Commission (2005).

Other data limitations are mentioned in Pickle (2002). In spatial research, many data limitations arise from the ability to track and identify patients from mortality data that are downloadable from Departments of Health. This makes it difficult to download data since this information must be by law, confidential. There are a few methods available to cover identities of patients which have recently been developed.

The identification of patients is evident in the United States as the data sets are small and thus the National Center For Health Statistics does not release mortality data for a county over a year, this information may be published over several years.

Another problem mentioned by Pickle (2002) is that covariate data may not be available or collected with the mortality data. Information about confounder variables such as lifestyle, family history and other variables are not collected and these variables may have a significant impact on the results of research. Environmental factors such as climate, the level of pollution, water quality and exposure to dangerous chemicals are also rarely used and not collected.

Another limitation is medical records of patients which may release more information about a disease but may not be used because of privacy and confidentiality.

3.2 Calculation of Expected Values

The expected values are calculated for the Georgia data based upon the statewide incidence rate. The standard population is selected as the population of Georgia as all counties in the study are a subset of the Georgia population. The population of each county is recorded along with the number of deaths. The total number of deaths in Georgia will be the sum of all the deaths in all of the counties. The

expected values will be calculated by:

$$e_i = m_i \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} m_i},$$

where m_i is the population in the i^{th} county. The mid-year population estimates by county for Georgia are given in Tables B.1 and B.2 in Appendix B.1.

The expected values for the South African data are calculated in the same way as the Georgia expected values are calculated. Both sets of expected values are calculated in RStudio®(2015). The South African data used for the calculation are given in Table 3.1.

Table 3.1: The mid-year population estimates in South Africa by province in 2014.

Province	**Population**
Eastern Cape	6786880
Free State	2786757
Gauteng	12914817
KwaZulu-Natal	10694434
Limpopo	5630464
Mpumalanga	4229323
North-West	3676274
Northern Cape	1166680
Western Cape	6116324
Total	**54001953**

3.3 Georgia Data

Georgia was founded as a British colony in 1733 which makes the state the youngest of the 13 British colonies, according to Reese (2010). The name Georgia was derived from George II who arrived on the land in 1733. The capital of Georgia is Atlanta. Georgia is most famous for its plantations and had the most number of plantations in the South by the 1950s.

The oral cancer for Georgia in 2004 by county is listed in Tables B.3 and B.4, the data can be found from the Georgia Department of Public Health Data Warehouse available at (`https://oasis.state.ga.us`). Figure 3.1 shows the observed deaths in each county. The expected values are calculated based upon the statewide incidence rate.

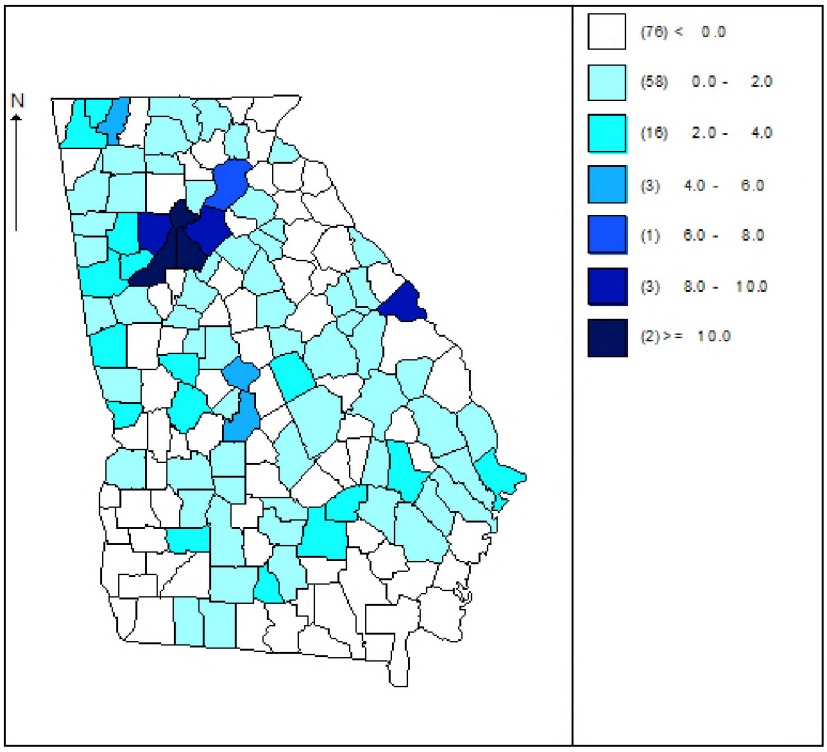


Figure 3.1: The observed number of deaths of oral cancer in Georgia in 2004.

3.3.1 Cancer

3.3.1.1 Overview

Cancer results in the second highest number of deaths worldwide after cardiovascular disease, according to Sudhakar (2009). Recent breakthroughs in treatment and technology have lead to a reduction in the number of deaths caused by cancer. Early identification is a major factor in the survival rate of a patient who has cancer as treatment methods may not be effective against cancer in later stages of tumours.

Cancer has been identified as early as 1600 B.C in Egypt. This was human bone cancer as stated by Sudhakar (2009). Other early records of cancer were found in manuscripts also in Egypt. There was no treatment available historically, surgery may have been performed for surface tumours but survival was still low after surgery.

Cancer is caused by the rapid growth of cells in a region of the body, and the growth becomes uncontrollable. Although cancer may be formed in many regions of the body and there are numerous types of cancer, the common characteristic is cells which proliferate, split and continue to split when healthy cells would die off and new cells would grow to replace the dead cells as written by Sudhakar (2009). Cancer may be spread through blood and veins and cells may travel from one region of the body to another region via the blood and start to grow in this region.

One of the causes of cancer is thought to be a damaged deoxyribonucleic acid (DNA) which should be repaired by the body but is not repaired, another cause may be due to the inheritance of damaged DNA through family or damaged caused by lifestyle as mentioned by Sudhakar (2009). Depending on the type of cancer, a solid tumour is often formed, but in some cancers, a tumour may not be formed. One example of this type of cancer is leukaemia, where cancer cells occur in the blood. Some tumours which grow are not cancerous and are referred to as noncancerous or a benignant tumour.

3.3.1.2 Treatment

The removal of cancer via surgery meant that sometimes cancer would be able to reoccur and tumours could grow again. The discovery of anaesthesia in 1846 resulted in a breakthrough in cancer treatment. Surgeons, named by Sudhakar (2009) as Bilroth, Handley and Halsted started to remove cancer tumours along with the lymph nodes. A surgeon named Paget, later discovered that cancer could be spread from a tumour to another part of the body through the blood, this is referred to as metastasis.

Methods to treat cancer were later developed and are named in Sudhakar (2009) as Chemotherapy, Hormonal therapy, Radiation therapy, Adjuvant therapy and Immunotherapy.

3.3.1.3 Oral Cancer

The term oral cancer used in this research refers to malignant neoplasms of lip, oral cavity and pharynx as specified by Georgia Department of Public Health Data Warehouse available at (`https://oasis.state.ga.us`). Oral cancer is defined by Kirita & Omura (2015) as a malignant neoplasm that grows in the oral region of a body. The oral region of the body is exposed to constant bacteria present in the environment, exposure due to smoking, pollution as well as chemicals present in food and alcohol. Other exposures named in Kirita & Omura (2015) are "mechanical stimuli" and are "ill-fitting prosthetic appliance" and are identified by several conditions which fall under carcinogenesis.

Early identification of cancer is vital in the treatment and survival of the cancer patient as outlined in Section 3.3.1.1. This identification is easier in oral cancer than in other cancers as identification of oral cancer can be made by a combination of palpation and optical observation. Oral cancer may be spread to other parts of the body which is also another reason for early detection.

3.3.2 Georgia Model

3.3.2.1 Model Parameters

The model with both UH and CH components is referred to as the convolution model, where the CH component is modelled using an ICAR prior and is used to model the Georgia data and is proposed by Lawson (2009). The number of iterations are set at $n_T = 50000$ and the burn-in period is set at $N_B = 12000$. Convergence is visually checked in the trace plots and the kernel densities of the

relative risks are analysed and model diagnostics are performed. The model fit for each model will be accessed using a quantile-quantile (Q-Q) plot of the deviance residuals. The deviance residuals of a model perform similarly to ordinary residuals in a "standard normal-theory linear regression model," as proposed by Montgomery et al. (2015). The SMR is smoothed by assigning a prior distribution to the calculation of the SMR values. A further $n_A = 20000$ iterations are used for each model to calculate the DIC. The code for the Georgia model is written by Lawson (2009) and modified to produce the following models.

3.3.2.2 Conjugate Prior

The Georgia model with a conjugate prior for the precision of the uncorrelated heterogeneity, denoted by v in Section *2.12.3.1*, is assigned a conjugate hyperprior distribution for the precision. This model will be referred to as model 1. The conjugate hyperprior distribution is a gamma distribution as the data is assumed to follow a Poisson distribution and the posterior distribution is also a gamma distribution. The parameters of the conjugate prior are derived from the data, this is done in Section 2.9.4, and resulted in parameters with values of $\alpha = 0.3139$ and $\beta = 0.2329$, respectively. The CH component is modelled by the ICAR prior which has a gamma hyperprior distribution for the precision with parameters with values of $\alpha = 0.1$ and $\beta = 0.0001$, respectively.

Table 3.2: An analysis of the burn-in period for model 1.

County	Percentage
Bryan	0.7544
Clinch	0.7688
Coweta	0.7419
Early	0.7856
Glynn	1.0949
Jackson	1.0953
Jenkins	1.4500
Liberty	0.8640
McIntosh	0.8004
Monroe	0.5864
Richmond	0.7264
Spalding	0.9011
Taliaferro	0.6182
Union	0.9019
Warren	0.9035
Wilkes	0.6282

A rule of thumb proposed by Woodward (2016) is that the burn-in period should be such that the Monte Carlo (MC) error is less than 5% of the standard deviation for a statistic. Approximately 10% of the counties in the model are randomly selected as the number of counties in the model are too

numerous to list in a table. These randomly selected counties are given in Table 3.2 and it may be seen that all the selected counties have an MC error which is less than 5% of the standard deviation for the relative risk and thus the burn-in period of $N_B = 12000$ may be considered as sufficient. Other results based on this model can be found in Appendix C.1.

The convergence of the model is assessed by the trace plots and kernel densities. Figures C.3 and C.4 show the trace plot and kernel densities for a sample of randomly selected counties as there are too many counties to include all of them in the results. These randomly selected counties will remain the same for the three models to make the comparison easier. The convergence for the model is easily seen in the trace plots in Figures C.3 and C.5 and the kernel densities in Figure C.4. The trace plot of the deviance in Figure C.5 also shows convergence.

Model fit was assessed using a Q-Q plot of the deviance residuals, Figure C.1, and a histogram of the Bayesian residuals, Figure C.2. The calculation of the Bayesian residuals is outlined in Section 2.13.4.3. Figure C.1 shows that the model fits the data adequately although there is evidence that the model has heavy tails. The histogram of the Bayesian residuals, Figure C.2, also shows that there is evidence that the Bayesian residuals have heavy tails. Other model fit statistics are given in Table 3.3. Selected results of the deviance residuals are given in Table C.1.

The calculation of the deviance residuals is given in Section 2.13.4.4. The deviance residuals are implemented in WinBUGS® by calculating the deviance contributed by each county using Equation 2.27 in Section 2.13.4.4. Equation 2.27 is implemented in WinBUGS® as: $Ds[i] < -2 * ((y[i] * log(y[i]/mu[i])) - (y[i] - mu[i]))$. The next step is to determine the sign in Equation 2.25 which requires the $step$ function in WinBUGS®. The $step$ function performs similarly to an if statement in other programming languages. The $step$ function returns a value of 1 if the result from the variables inserted into the function are greater than or equal to 0 and will return a value of 0 otherwise. The code to determine the sign is: $sign[i] < -2 * step(y[i] - mu[i]) - 1$. The $step$ function will thus return a value of 1 if the result of $y[i] - mu[i]$ is greater than or equal to 0 and will return a value of 0 otherwise. The result of the $step$ function will then be multiplied by 2. Thus if the $step$ function returned a value of 1, it would be multiplied by 2 to give 2 and if the $step$ function returned a value of 0, it would be multiplied by 2 to give 0. The last part is to subtract 1, then if we have a value of 2 and subtract 1, the result will be a 1 and if we have a value of 0 and subtract 1 the result will be -1. This then completes the sign part of Equation 2.25. The deviance residual contributed by each county is thus calculated by the following code: $dev.res[i] < -sign[i] * sqrt(Ds[i])$. The code for the calculation of the deviance residuals is written by Lunn et al. (2012).

Table 3.3: The model diagnostics of model 1.

Statistic	Mean	Standard Deviation	Monte Carlo Error	95% Confidence Interval	
				2.5%	97.5%
MAPE	0.9933	0.0869	8.514E-4	0.8302	1.17
MSPE	2.752	0.701	0.004819	1.736	4.465
Deviance	378.1	15.52	0.3272	348.3	408.7
pE	0.7719	0.06582	0.001917	0.6381	0.8941
pV	0.2016	0.07488	0.003093	0.06286	0.3495
pS	0.02651	0.04703	0.002974	3.614E-5	0.163
Moran's I	0.006929	0.03249	7.83E-4	-0.06331	0.06427
DIC	416.772				
pD	38.704				
Dbar	378.068				
Dhat	339.364				

The model diagnostics of model 1 are given in Table 3.3 and will be used to make comparisons with the other two models for the Georgia data. The MAPE and MSPE are calculated for all three models and the method of the calculation is explained in Section 2.13.2.

The calculation of the MAPE and MSPE statistics are easily accomplished in WinBUGS®. The predicted value, denoted by y^*_{ij}, is determined by assuming that the predictive values have a Poisson distribution and thus: $y^*_{ij} \sim Poiss\big(e_i\theta_i\big)$. In WinBUGS®, this is accomplished by: $ypred[i] \sim dpois(\mu[i])$. The numerator of the equation used to calculate the MAPE is $|y - y^*_{ij}|$ which is implemented in the model as: $PPL2[i] < -abs(ypred[i] - y[i])$, where the function abs results in the absolute value. The MAPE is then calculated by averaging the sum of $PPL2$ as: $MAPE = mean(PPL2[])$. The MSPE is calculated using the pow function as: $PPL[i] < -pow(ypred[i] - y[i], 2)$ which is the numerator, $\big(y_i - y^*_{ij}\big)^2$, in the equation for the MSPE. The MSPE is then obtained by averaging the sum of PPL, as: $MSPE < -mean(PPL[])$.

The statistics, denoted by pE, pV and pS, represent the proportion of the total variation in the model due to the expected values, unobserved non-spatial factors and unobserved spatial factors, respectively. The calculations of these statistics are based on the method of Lunn et al. (2012). The calculation requires the log of each expected value to be calculated, this is done by: $Lexp[i] < -log(e[i])$, where $e[i]$ is the expected value of the i^{th} county. This code needs to be executed within the model $Loop$. The standard deviation of the log expected value, UH component and CH component are then calculated by: $sdE < -sd(Lexp[])$, $sdV < -sd(v[])$ and $sdS < -sd(S[])$, respectively. The sum of each of the variances is required to calculate the proportions, the sum is: $sum < -sdS*sdS+sdV*sdV+sdE*sdE$. Each proportion is then calculated by dividing the variance of each statistic by the sum of the variances of all the statistics. The code required is: $pS < -sdS * sdS/sum$, $pV < -sdV * sdV/sum$, and $pE < -sdE * sdE/sum$ for the proportion of the unobserved spatial factors, unobserved non-spatial factors and variation due to the expected values, respectively. There is 77.19% of the total variation in

the model due to the expected values and is represented by pE. The proportion of the total variation in the model due to the unobserved non-spatial factors, pV, is 20.16% and the proportion of the total variation in the model due to unobserved spatial factors, pS, is 2.651%.

The deviance is calculated using the built in $deviance$ function in WinBUGS® and is calculated as 378.1.

The DIC is also calculated by a built in function in WinBUGS® and based on a further $n_A = 20000$ iterations after $n_T = 50000$ iterations are run. The DIC is 416.772.

The posterior average estimate of the Moran's Autocorrelation Coefficient has been calculated for the Georgia models. The calculation of the Moran's Autocorrelation Coefficient is given in Section 2.4.2. The calculation of the posterior average estimate of the Moran's Autocorrelation Coefficient requires a significant amount of computations and there is a high computation time associated with the calculation. Since there is a large amount of computation involved in the calculation, we have performed some of the calculation in RStudio® and the rest of the calculation in WinBUGS®. The first part of the calculation, which is done in RStudio®, is to calculate the cumulative sum of the neighbours of all of the counties, with the first neighbour of the cumulative sum given the value of 0, $csum[1] < -0$. The cumulative sum is the cumulative sum of the number of neighbours vector. The number of neighbours, denoted $\boldsymbol{num}[]$, is produced using the $adjacency\ tool$ in GeoBUGS®. The cumulative sum of the neighbours is used to select the elements of the residuals for the calculation. The code for the cumulative sum is:

$$\begin{aligned} csum &< - rep(0, 160) \\ csum[1] &< - 0 \\ for(i\ in\ 2 : (159 + 1)) &\ \{ \\ csum[i] &< - sum(num[1 : (i - 1)])\}. \end{aligned}$$

The $loop$ starts at 2 since the first neighbour of the cumulative sum is given the value of 0. The next step, which is also done in RStudio®, is to determine which elements of the neighbouring residuals will be selected. The code to determine the index of the elements which will be selected is as follows:

$$\begin{aligned} x &< - rep(0, 159) \\ g &< - rep(0, 159) \\ for(i\ in\ 1 : 159) &\ \{ \\ x[i] &< - csum[i] + 1 \\ g[i] &< - csum[i + 1]\}, \end{aligned}$$

where $x[i]$ is the lower index of the element which will be selected for the i^{th} county and $g[i]$ is the upper index of the element which will be selected for the i^{th} county.

The results produced in RStudio® are then inserted into WinBUGS® as data, these results are the cumulative sum of neighbours and the lower and upper indexes for the elements of the neighbouring residuals that will be selected.

The first step in the WinBUGS® part of the calculation is to determine the standardised Bayesian residuals, this is done by: $res[i] < -(y[i] - mu[i])/sqrt(mu[i])$. The next step is to produce the adjacency matrix, $We[]$, of the residuals for the neighbouring counties. The code is:

$$\begin{aligned} for(k\ in\ 1 : sumNumNeigh) \quad & \{ \\ We[k] \ <- \ & res[adj[k]]\}, \end{aligned}$$

where $sumNumNeigh$ is the sum of the number of neighbouring counties, $sum(\boldsymbol{num}[])$. The sum of the neighbouring residuals is then calculated by: $estar[i] < -sum(We[x[i] : g[i]])$, where $\boldsymbol{x}[]$ and $\boldsymbol{g}[]$ are inserted as data and calculated in RStudio®. The difference between the i^{th} residual value and the mean residual value is then calculated as: $de[i] < -res[i] - mean(res[])$ and the difference between the i^{th} neighbouring residual and the mean of the neighbouring residuals is: $d.estar[i] < -estar[i] - mean(estar[])$. The numerator of Equation 2.2 is calculated as: $dt[i] < -de[i] * d.estar[i]$ and the denominator of Equation 2.2 is calculated as: $db[i] < -pow(d.estar[i], 2)$ and the estimate of the Moran's Autocorrelation Coefficient is then the numerator divided by the denominator: $Moran < -sum(dt[])/sum(db[])$. The estimate of the Moran's Autocorrelation Coefficient is 0.006929 which indicates that there is little autocorrelation left in the model after the model has been fitted.

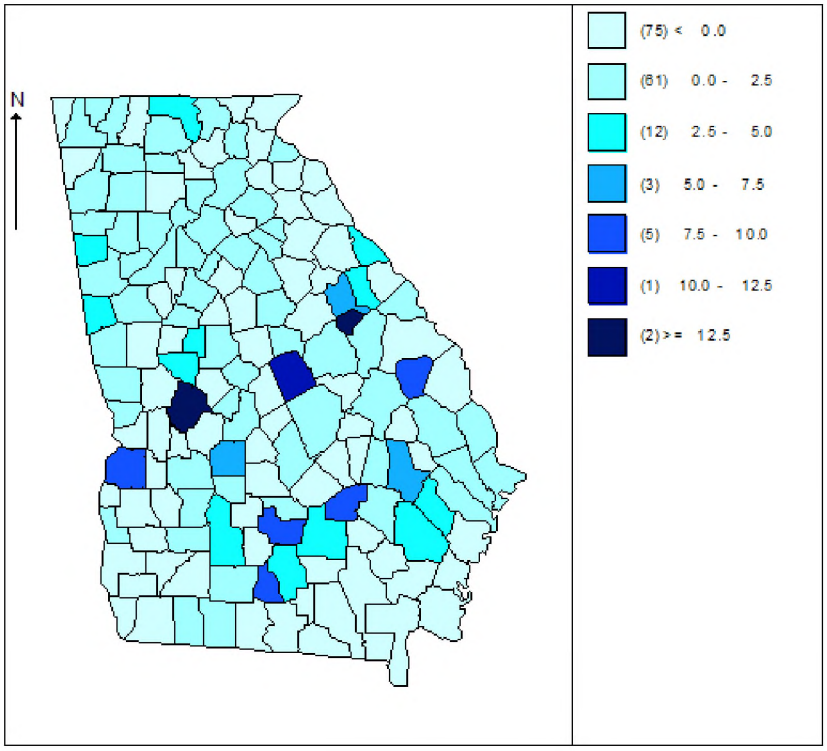


Figure 3.2: The SMR for model 1.

The SMR in Figure 3.2 has been smoothed by assigning a prior distribution to the calculation of the SMR. The smoothing is proposed by Lawson (2009) in which a prior distribution, denoted by $eps2$, is included in the calculation of SMR. The code for the calculation is: $smr[i] < -(y[i] + eps2)/(e[i] + eps2)$ where the prior distribution is assigned as follows: $eps2 \sim dnorm(0, 1000)$. This prior results in very small values being added and subtracted to the numerator and denominator of the SMR calculation. Figure 3.2 shows that there are eight counties which experience relatively high SMR values and are listed in Table 3.4.

Table 3.4: The selected results of the SMR of model 1.

County	SMR			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Glascock	12.550	2.258	0.005733	10.374	16.740
Taylor	13.790	1.960	0.009412	10.780	18.360
Wilkinson	12.100	1.473	0.007073	9.758	15.480
Stewart	8.713	9.210	0.045610	5.733	15.310
Irwin	8.276	0.994	0.004769	6.706	10.560
Cook	7.515	0.525	0.002519	6.606	8.654
Jeff Davis	9.563	0.885	0.004249	8.086	11.530
Jenkins	9.524	1.359	0.006531	7.458	12.710

The highest SMR values occur at Glascock and Taylor with posterior expected SMR values of 12.550 and 13.790, respectively. The high SMR value for Glascock is a result of a lower expected number of deaths of 0.065 and this combined with 1 death in the county result in a high SMR value. Taylor has 3 observed deaths and an expected value of 0.2218. This low expected value relative to a high number of deaths results in a high SMR value. This is one of the disadvantages of using the SMR as mentioned in Section 2.5.3.2.

Another disadvantage of using the SMR is that counties with a 0 observed number of deaths result in a SMR value of 0 (without smoothing), irrespective of the value of the expected number of deaths and is illustrated in the disease map in Figure 3.2.

Counties with relatively high observed values and a small expected value result in a large SMR value.

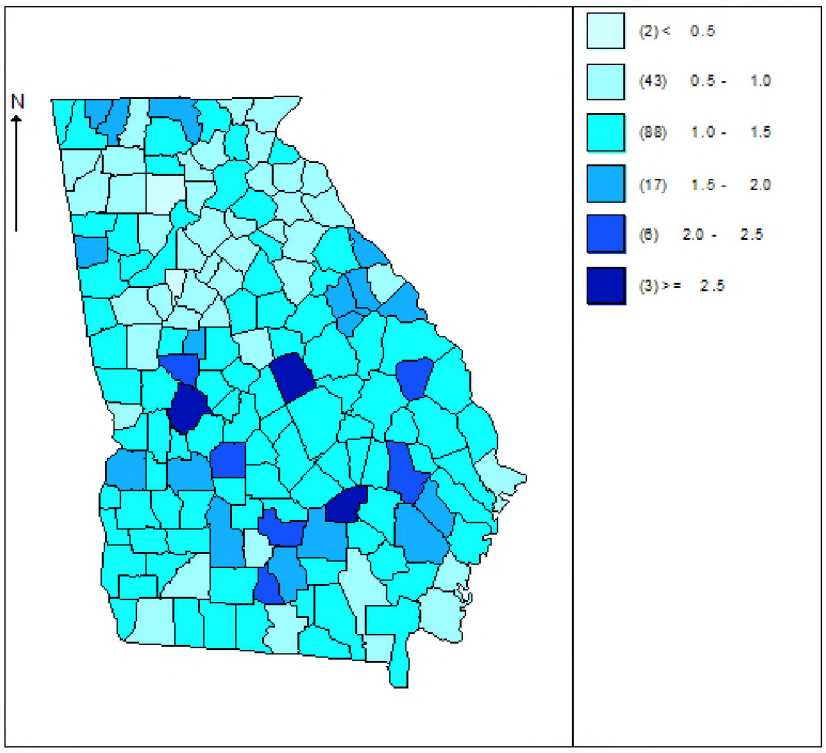


Figure 3.3: The relative risk for model 1.

The "null risk value," which is a relative risk value of 1, is usually the value in which there is no difference in the experimental and control groups. In this case, there would be no difference between the study region and the standard population. A relative risk value which is less than one indicates that there is a lower probability of the event of interest occurring in the study county than what we would expect from the standard population and a relative risk value which is higher than 1 indicates that there is a higher probability of the event of interest occurring in the study county than what we would expect from the standard population. There are 88 counties which are between 1-1.5 which means that there is a higher probability of deaths occurring in these counties than what would be expected from the standard population. Some of the 88 counties may have a relative risk which is marginally higher than 1 or equal to 1 and thus would not be significant enough to indicate a high risk of contracting and dying from oral cancer. There are 45 counties which have a relative risk which is less than 1 and we would expect that the probability of a death occurring in these counties is lower than we would expect from the standard population. There are 9 counties which have a relative risk more than or equal to 2, these counties are given in Table 3.5. It would be expected that there is a higher probability of a death taking place in these counties than in the standard population.

Table 3.5: The selected results of the relative risk of model 1.

County	Relative Risk			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Wilkinson	2.839	1.871	0.02259	0.8008	7.772
Jeff Davis	2.705	1.704	0.01946	0.7812	7.131
Taylor	2.934	2.010	0.02557	0.8113	8.271
Jenkins	2.246	1.552	0.01777	0.5931	6.394
Tattnall	2.235	1.244	0.01270	0.7088	5.393
Irwin	2.187	1.450	0.01693	0.5886	6.005
Cook	2.498	1.519	0.01626	0.7353	6.413
Dooly	2.091	1.335	0.01404	0.5848	5.571
Upson	2.030	1.091	0.01059	0.6653	4.848

Taylor has the highest relative risk with a relative risk of 2.934, this follows from the results of the SMR. The relative risk is also higher than 1 which would indicate that there are more deaths in the county than we would expect from the standard population. All of the counties in Table 3.5 have relative risks which are higher than 1 and thus the same can be concluded.

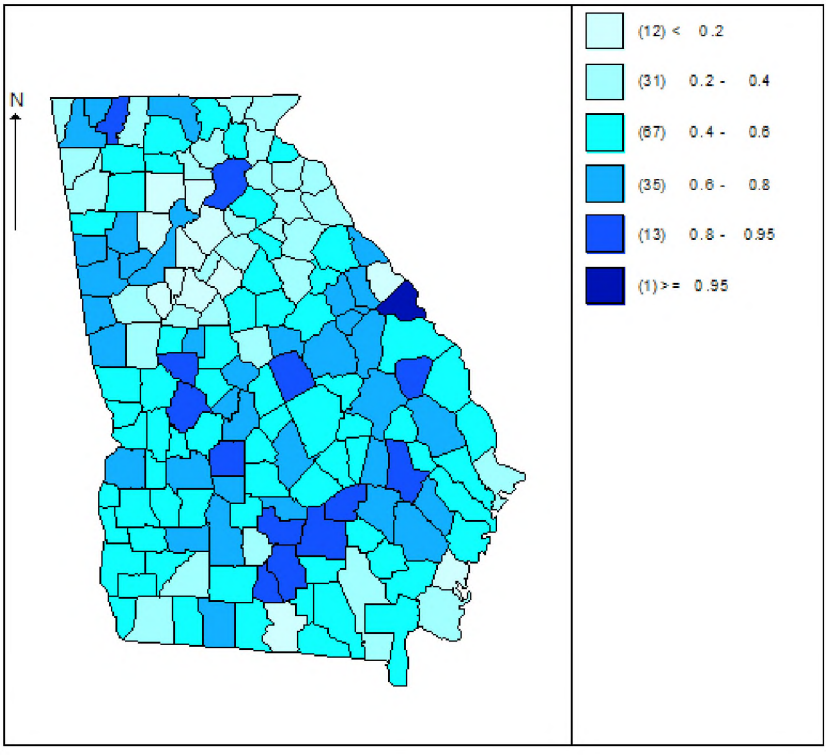


Figure 3.4: The posterior expected exceedance probability for model 1.

The calculation of the PP is outlined in Section 2.13.3. The code to calculate the PP involves using the *step* function in WinBUGS®. The calculation of the PP is as follows: $PP[i] < -step(theta[i]-1+$

$eps)$, where $theta[i]$ is the relative risk in the i^{th} county and $eps < -0.000001$ is the prior distribution. The $step$ function will return a value of 1 if $theta[i] - 1 + eps \geq 0$ and will return a value of 0 when $theta[i] - 1 + eps < 0$. The PP is then averaged over the number of samples of the model.

The PP gives an indication of which counties would be expected to have unusual or unexpected clusters in the disease. The disease map of the PP follows from the disease map of the relative risk in that the counties with high relative risk also have a high PP value. There are 12 counties which have PP values which are less than 0.2. Most of the counties show that there is little evidence of clustering in the Georgia data for model 1. There is 1 county which has a PP value greater than or equal to 0.95. Counties which have high PP values are given in Table 3.6.

Table 3.6: The selected results of the posterior expected exceedance probability of model 1.

Posterior Expected Exceedance Probability			
County	**Mean**	**Standard Deviation**	**MC Error**
Richmond	0.9516	0.2145	0.001118
Jenkins	0.8537	0.3534	0.002379
Tattnall	0.9021	0.2972	0.002074
Jeff Davis	0.9331	0.2499	0.001806
Coffee	0.8476	0.3594	0.002523
Irwin	0.8526	0.3545	0.002862
Berrien	0.8137	0.3894	0.002679
Cook	0.9174	0.2753	0.001860
Dooly	0.8432	0.3636	0.002803
Taylor	0.9418	0.2341	0.001536
Upson	0.8745	0.3313	0.001973
Whitfield	0.8616	0.3453	0.002176
Wilkinson	0.9376	0.2419	0.001684

The highest PP value for model 1 is the Richmond County which has a PP value of 0.9516, this may indicate that there is unusual or unexpected clustering in the county. The confidence intervals for the PP are not given in Table 3.6 as all of the confidence intervals have a lower bound of 0 and an upper bound of 1.

Using the notation from Section 2.13.3, let the threshold level be $c = 1$ and the exceedance probability of interest be $\hat{P} = 0.95$. There is only 1 county which exceeds the exceedance probability of interest, 0.95, which is Richmond and thus there may be evidence of unusual or unexpected clustering occurring in this county although further investigation is necessary to prove this and is not covered in this thesis. The PP value of Richmond only marginally exceeds 0.95 so evidence for clustering may be insignificant but still needs further investigation.

3.3.2.3 Jeffreys' Prior

This model will be referred to as model 2. This model also has the CH component modelled by the ICAR prior with a gamma hyperprior distribution for the precision with parameters of $\alpha = 0.1$ and $\beta = 0.0001$, respectively, the same parameters as model 1. The hyperprior distribution for the precision of the UH component is modelled by a gamma distribution with parameters of $\alpha = 0.5$ and $\beta = 0.0001$, respectively, which are the parameters of the Jeffreys' prior.

Table 3.7: An analysis of the burn-in period for model 2.

County	Percentage
Bryan	0.7981
Clinch	0.8742
Coweta	1.3595
Early	1.1788
Glynn	1.4469
Jackson	1.6961
Jenkins	1.3726
Liberty	0.9528
McIntosh	0.8075
Monroe	0.7489
Richmond	0.9866
Spalding	1.1233
Taliaferro	0.7990
Union	0.9316
Warren	1.0775
Wilkes	0.8325

Table 3.7 shows the proportion of the MC error to the standard deviation for the relative risk of model 2. All of the proportions are well below 5% and thus the burn-in period is adequate at $N_B = 12000$.

The convergence of model 2 will be assessed by trace plots and kernel densities, as was the case for model 1. The trace plots and kernel densities for model 2 are given in Figures C.8 and C.9 in Appendix C.2. The trace plots and kernel densities show that the model does converge. The model fit is assessed by the Q-Q plot of the deviance residuals and a histogram of the Bayesian residuals. Other model fit statistics are given in Table 3.8. The Q-Q plot, Figure C.6, shows that the model fit is adequate although there is evidence of heavy tails, as in model 1. The histogram of the Bayesian residuals also shows the heavy tails which can be seen in Figure C.7.

The DIC will be compared with the other models to determine which model is the better model to use when analysing the Georgia data.

Table 3.8: The model diagnostics of model 2.

Statistic	Mean	Standard Deviation	Monte Carlo Error	95% Confidence Interval	
				2.5%	97.5%
MAPE	1.021	0.08917	0.001344	0.8553	1.201
MSPE	2.843	0.6853	0.006855	1.811	4.459
Deviance	386.7	15.63	0.5154	355.5	416.8
pE	0.8092	0.06853	0.002589	0.6679	0.9306
pV	0.05832	0.07442	0.00455	6.795E-5	0.2573
pS	0.1325	0.07661	0.004055	0.001683	0.2905
Moran's I	-0.01164	0.0342	0.001063	-0.08282	0.05126
DIC	419.436				
pD	32.046				
Dbar	355.343				
Dhat	387.390				

The model diagnostics of model 2 are given in Table 3.8. The proportion of unobserved non-spatial factors, pV, is 5.832% of the total variation in the model which is 14.328% lower than model 1. This is the result of changing the parameters of the hyperprior distribution of the precision of the UH component in model 2. The proportion of unobserved spatial factors, pS, is 13.25% of the total variation in the model, compared with 2.651% in model 1. This is a difference of 10.599% which shows that the CH component is dominant over the UH component in model 2, whereas in model 1 the UH component was dominant over the CH component.

The MAPE and MSPE are 1.021 and 2.843, respectively and in model 1 the MAPE and MSPE are 0.9933 and 2.752, respectively. This is a difference of 0.0277 and 0.091 for the MAPE and MSPE, respectively. The MAPE and MSPE are both lower in model 1 than in model 2 which indicates that model 1 is better than model 2.

The DIC of model 1, 416.772, is lower than that of model 2, 419.436. The difference in DIC of 2.664 is not significant as this difference is less than 10. The deviance of model 1 is also lower than that of model 2 by a difference of 8.6 which indicates that model 1 is the better model. The posterior estimate of the Moran's Autocorrelation Coefficient is -0.01164 which indicates that there is little autocorrelation left in the model after the model has been fitted.

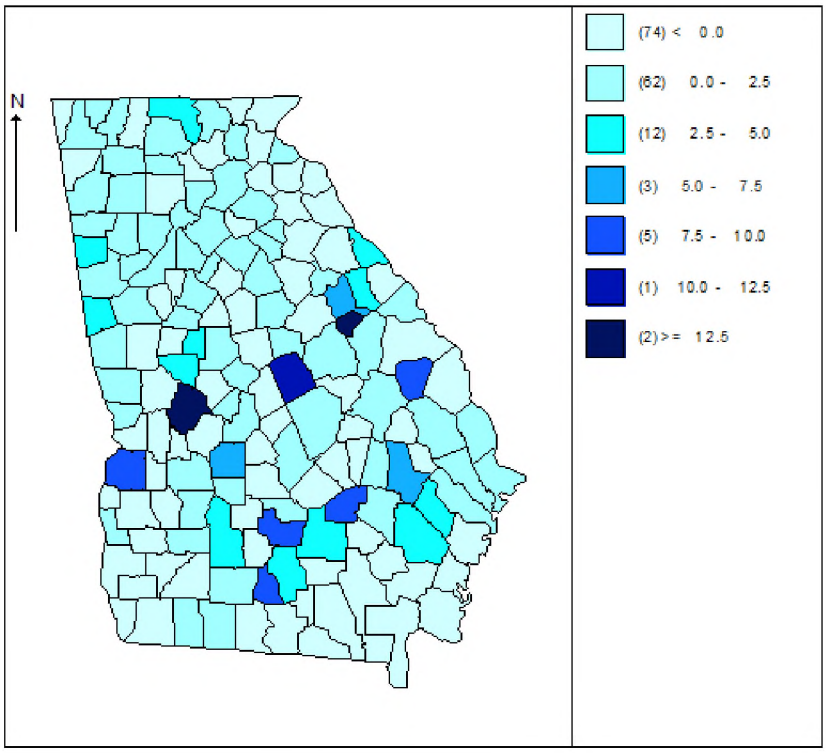


Figure 3.5: The SMR for model 2.

The disease map for the SMR of model 2 displays similar results to those of model 1. The only difference between the two disease maps is that in model 1 there are 75 counties which have a SMR of 0 whereas in model 2 there are 74. The other difference is that model 2 has 62 counties in category 0.0-2.5 whereas model 1 has 61. There are 8 counties which have SMR values which are greater than 7.5, these counties are given in Table 3.9.

Table 3.9: The selected results of the SMR of model 2.

County	SMR			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Glascock	12.510	1.642	0.007308	10.420	15.350
Taylor	13.780	1.949	0.010650	10.790	18.270
Wilkinson	12.100	1.466	0.007982	9.764	15.410
Stewart	8.774	17.790	0.090030	5.739	15.120
Irwin	8.272	0.989	0.005385	6.710	10.520
Cook	7.513	0.522	0.002824	6.609	8.635
Jeff Davis	9.560	0.881	0.004778	8.090	11.500
Jenkins	9.518	1.352	0.007391	7.463	12.640

The highest SMR value occurs in Taylor with a value of 13.780, this follows from the results of model 1. Comparing Table 3.9 with Table 3.4, the SMR results from model 1, most of the mean SMR

values in Table 3.9 are higher than those of Table 3.4, with Stewart being the only county in the table which is lower in model 1 than in model 2.

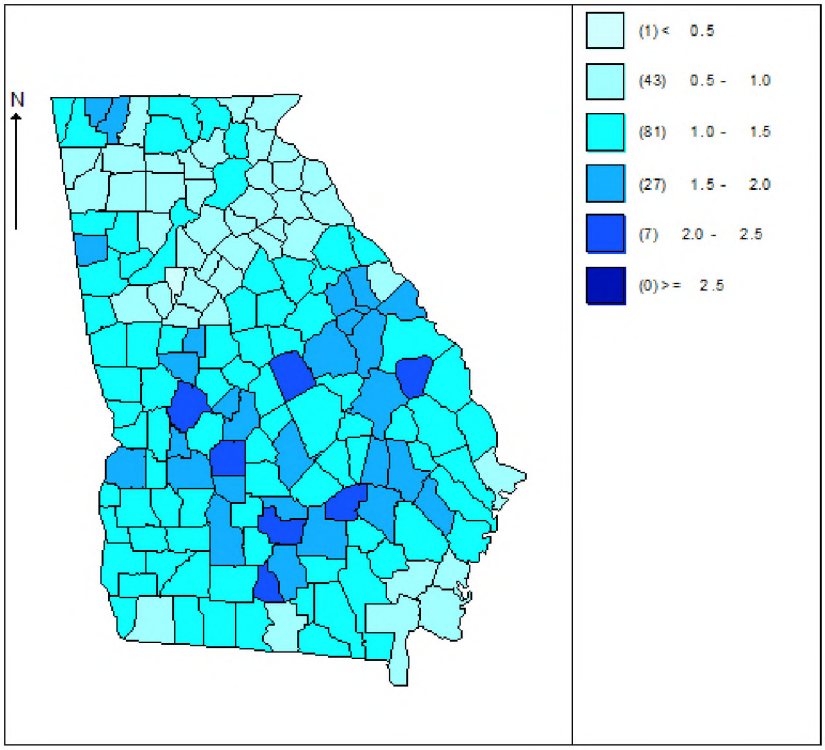


Figure 3.6: The relative risk for model 2.

The relative risk of the model only has 1 county which is less than 0.5 and a further 43 which are between 0.5 and 1. This would indicate that there is a low probability of deaths occurring in these counties than we would expect from the standard population. There are 81 counties who have a relative risk between 1 and 1.5 whereas in model 1 there were 88 counties between 1 and 1.5. High relative risk values occur at 7 counties, these are listed in Table 3.10 along with 2 other counties which had high relative risk values for model 1.

Table 3.10: The selected results of the relative risk of model 2.

County	Relative Risk			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Wilkinson	2.261	1.301	0.02777	0.8489	5.6740
Jeff Davis	2.273	1.236	0.02373	0.8515	5.5310
Taylor	2.358	1.418	0.02949	0.8606	6.0370
Jenkins	2.108	1.299	0.01783	0.6629	5.5460
Irwin	2.161	1.232	0.01776	0.7389	5.4090
Cook	2.087	1.167	0.02087	0.7271	5.1550
Dooly	2.032	1.074	0.01547	0.7461	4.7950
Tattnall	1.932	0.928	0.01627	0.7856	4.3120
Upson	1.773	0.839	0.01291	0.7188	3.9370

The highest relative risk value is at Taylor for model 2, this was also the highest relative risk value for model 1. The difference in relative risk for Taylor in the 2 models is 0.576, with model 1 having a higher relative risk for Taylor with a value of 2.934. The relative risk values in Table 3.10 are all lower than those from model 1 in Table 3.5. The biggest difference between the 2 tables is in Wilkinson which has a difference of 0.578 between the two models. The standard deviation in all the counties listed in Table 3.10 are lower than those of model 1 in Table 3.5.

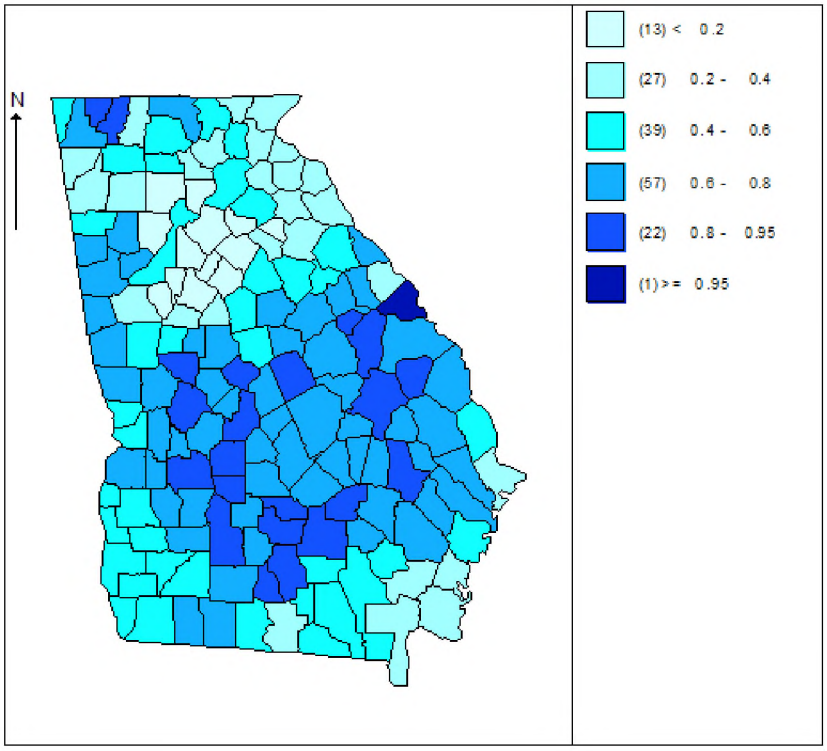


Figure 3.7: The posterior expected exceedance probability for model 2.

There is only 1 county which has a PP value higher than or equal to 0.95, this is the same as model 1. The other categories in model 2 show a different number of counties than what is shown in model 1. The category for the PP less than 0.2 has 13 counties for model 2 where model 1 has 12, a difference of 1. The next category, with PP values between 0.2 and 0.4, has 27 counties for model 2 and 31 counties for model 1 which is a higher difference than the previous category. The third category, 0.4-0.6, has 39 counties for model 2 compared with 67 for model 1, a difference of 28. Other categories with high differences between model 2 and 1 are the 0.6-0.8 category which has 57 counties for model 2 compared with 35 for model 1, a difference of 22.

Table 3.11: The selected results of the posterior expected exceedance probability of model 2.

Posterior Expected Exceedance Probability			
County	**Mean**	**Standard Deviation**	**MC Error**
Richmond	0.9500	0.2179	0.001312
Jenkins	0.8778	0.3276	0.002370
Tattnall	0.9138	0.2807	0.002134
Jeff Davis	0.9426	0.2326	0.001866
Coffee	0.9019	0.2974	0.002699
Irwin	0.9097	0.2866	0.002659
Berrien	0.8507	0.3564	0.002838
Cook	0.8982	0.3023	0.002270
Dooly	0.9080	0.2890	0.002697
Taylor	0.9451	0.2277	0.001628
Upson	0.8763	0.3292	0.002244
Wilkinson	0.9403	0.2369	0.001803
Whitfield	0.8495	0.3575	0.002801

Let the threshold level be $c = 1$ and the exceedance probability of interest be $\hat{P} = 0.95$. There is only 1 county which is equal to the exceedance probability of interest, 0.95, which is Richmond and thus there is no evidence of unusual or unexpected clustering in the model as the PP of Richmond does not exceed the exceedance probability of interest. Further investigation into clustering is required as further simulations may result in the PP of Richmond exceeding the exceedance probability of interest.

3.3.2.4 Fixed Parameter Prior

The model with a fixed parameter for the UH component prior distribution will be called model 3. The code for this model is written by Lawson (2009). The CH component which is modelled by the ICAR prior has a hyperprior distribution for the precision with parameters $\alpha = 0.005$ and $\beta = 0.005$, respectively. The UH component will be modelled with a zero mean Gaussian distribution with a precision parameter of 0.000001. This choice is common in research although it is a poor choice for the Georgia data as will be shown in this Section.

Table 3.12: An analysis of the burn-in period for model 3.

County	Percentage
Bryan	0.5385
Clinch	0.5198
Coweta	0.4995
Early	0.5153
Glynn	0.4839
Jackson	3.2868
Jenkins	3.2139
Liberty	3.0505
McIntosh	0.5128
Monroe	0.5238
Richmond	0.6205
Spalding	0.4905
Taliaferro	0.4946
Union	0.5465
Warren	0.5019
Wilkes	0.5123

The analysis of the burn-in period, which is set as $N_B = 12000$, shows that since all of the proportions of the MC error to the standard deviation of the relative risks are well below 5%, the burn-in period is sufficient. The convergence of model 3 is very poor for some of the counties. The trace plots for the relative risks of the selected counties are given in Figures C.13 and C.14. The trace plots in Figure C.13 show very poor convergence for counties Clinch, Early and Glynn and for Figure C.14 poor convergence is shown in counties McIntosh, Taliaferro and Wilkes. These are only the counties which were randomly selected, other counties in the model also have poor convergence.

The kernel densities in Figure C.15 show poor shapes for the counties which did not show convergence in the trace plots. The kernel density of the deviance also shows a poor shape and the trace plot of the deviance shows poor convergence. The Q-Q plot for the deviance residuals for model 3 are shown in Figure C.11 and the histogram of the Bayesian residuals shown in Figure C.12 show that the model fit is poor. The Bayesian residual histogram shows that the Bayesian residuals have a heavy negative skew. Thus we may conclude based on the trace plots, kernel densities, Q-Q plot and the histogram of the Bayesian residuals that the model does not fit the data adequately and convergence is only reached for some of the counties and not all counties in the model. The model will still be analysed and compared to models 1 and 2 and used only for illustrative purposes.

The model diagnostics are shown in Table 3.13.

Table 3.13: The model diagnostics of model 3.

Statistic	Mean	Standard Deviation	Monte Carlo Error	95% Confidence Interval	
				2.5%	97.5%
MAPE	1.048	1.572	0.03634	0.6604	2.836
MSPE	2.952	2.625	0.04349	1.547	5.717
Deviance	836.6	4495.0	53.53	286.4	3689.0
pE	4.089E-6	6.955E-7	4.082E-9	2.942E-6	5.664E-6
pV	1.0	2.911E-5	1.927E-6	0.9999	1.0
pS	9.731E-6	2.907E-5	1.927E-6	6.336E-9	1.078E-4
Moran's I	-0.007194	0.04278	3.902E-4	-0.09282	0.0766
DIC	448.447				
pD	110.120				
Dbar	338.326				
Dhat	228.206				

The most notable difference between model 3 and models 1 and 2 is the proportion of the total variation due to the unobserved non-spatial factors, pV, which is 100% for model 3. This is the result of the fixed parameter for the prior distribution of the UH component. Models 1 and 2 have a gamma hyperprior distribution which allows the precision to be estimated by the model and vary while model 3 has a fixed value and thus the precision will not vary and cannot be estimated. The fixed value of the precision may be changed to reduce the proportion of the variation total due to unobserved non-spatial factors. The high proportion of the total variation due to unobserved non-spatial factors means that the proportions of the total variation due to unobserved spatial factors and due to the expected values are both reduced. The proportion of the total variation due to the unobserved spatial factors, pS, is 0.0009731% compared to 2.651% of model 1 and 13.25% of model 2, differences of 2.65% and 13.25%, respectively. The proportion of the total variation due to the expected values is only 0.0004089% whereas model 1 has 77.19% and model 2 has 80.92%, differences of 77.19% and 80.92%, respectively.

The MAPE and MSPE in model 3 are both higher than those of models 1 and 2 which indicates that models 1 and 2 are the better models. This is expected as the model fit of model 3 to the Georgia data is very poor. Another indication of poor fit is provided by the deviance, in model 3 the deviance is 836.6 which compares to 378.1 of model 1 and 386.7 of model 2. The standard deviation of the deviance of model 3 is also significantly higher than the standard deviation of the deviance for models 1 and 2.

The DIC of model 3 is higher than models 1 and 2 which is to be expected and thus model 3 is the worst model out of the 3 models. Model 1 is thus the best model out of the 3 models.

The estimate of the Moran's Autocorrelation Coefficient is -0.007194 which indicates that there is little autocorrelation left in the model after the model has been fitted. The low autocorrelation follows from the results of models 1 and 2.

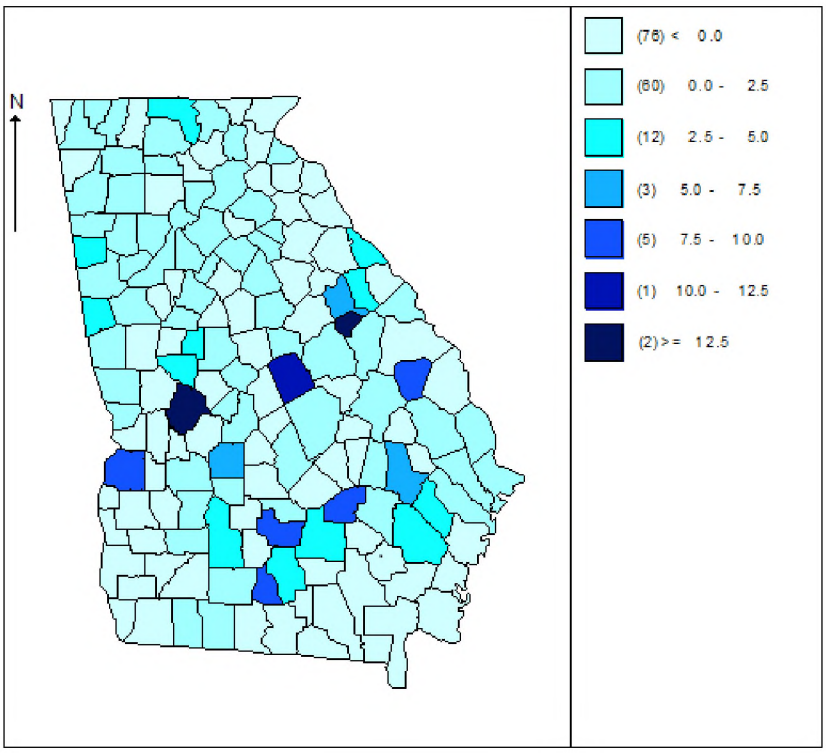


Figure 3.8: The SMR for model 3.

The disease map of the SMR of model 3 has a few differences from that of models 1 and 2. The differences in the disease maps are small and are due to the nature of the simulation. The similar results in the SMR values of all 3 models are expected since the SMR is calculated from the data and all 3 models are run off the same data. The prior distribution of the SMR results in small differences in the SMR values across the 3 models.

Table 3.14: The selected results of the SMR of model 3.

County	SMR			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Glascock	12.780	1.780	0.008803	9.850	16.580
Taylor	13.800	1.979	0.011010	10.800	18.390
Wilkinson	12.110	1.484	0.008257	9.771	15.500
Stewart	8.778	5.340	0.028500	5.744	15.380
Irwin	8.281	1.001	0.005570	6.715	10.580
Cook	7.518	0.526	0.002932	6.612	8.660
Jeff Davis	9.568	0.889	0.004951	8.095	11.550
Jenkins	9.532	1.374	0.007645	7.469	12.730

The SMR values for model 3 in Table 3.14 are similar to those from models 1 and 2. The highest SMR in model 3 occurs at Taylor which is consistent with models 1 and 2.

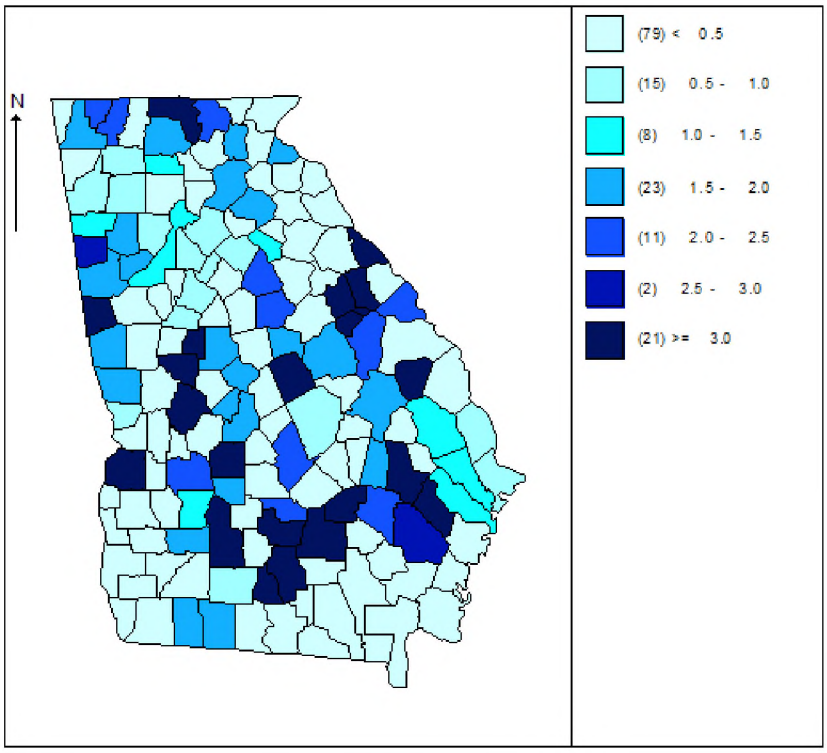


Figure 3.9: The relative risk for model 3.

The relative risk for model 3 is different to the relative risk of both models 1 and 2. The difference is due to the UH component because the relative risk incorporates the UH component. The relative risk for models 1 and 2 seems to be scaled down whereas in model 3, the scaling does not happen and thus the relative risks have very high values. The scaling of the relative risks in models 1 and 2 makes the comparison of each relative risk to the "null risk value" of 1, easier. The high relative risks of model 3 are far away from 1 and possibly indicate very high probability of more deaths occurring in these counties than we would expect from the standard population. This results in a false indication of a high probability of more deaths when it is not the case. The counties with high relative risk values do mostly follow from those of models 1 and 2, this is apparent in Table 3.15 which shows the same counties as models 1 and 2.

Table 3.15: The selected results of the relative risk of model 3.

County	Relative Risk			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Taylor	13.670	8.169	0.045510	2.719	33.600
Wilkinson	11.990	7.167	0.036640	2.463	29.460
Jenkins	9.422	6.831	0.036060	1.122	26.840
Dooly	6.998	5.041	0.024390	0.832	19.710
Irwin	8.248	6.022	0.034850	0.993	23.390
Cook	7.514	4.521	0.024830	1.508	18.580
Jeff Davis	9.547	5.660	0.028930	1.920	23.390
Glascock	15.510	15.830	0.082830	0.393	57.950
Stewart	8.136	8.261	0.041950	0.194	30.250
Tattnall	5.318	3.152	0.017760	1.063	12.940
Upson	4.360	2.610	0.014340	0.871	10.820

Table 3.15 shows the selected results of the relative risks of model 3 and follows from those of models 1 and 2 in which the same counties appear. The relative risks are much higher in model 3 than in models 1 and 2 and mostly follow from the SMR values. It is difficult to draw conclusions based on the high relative risk values.

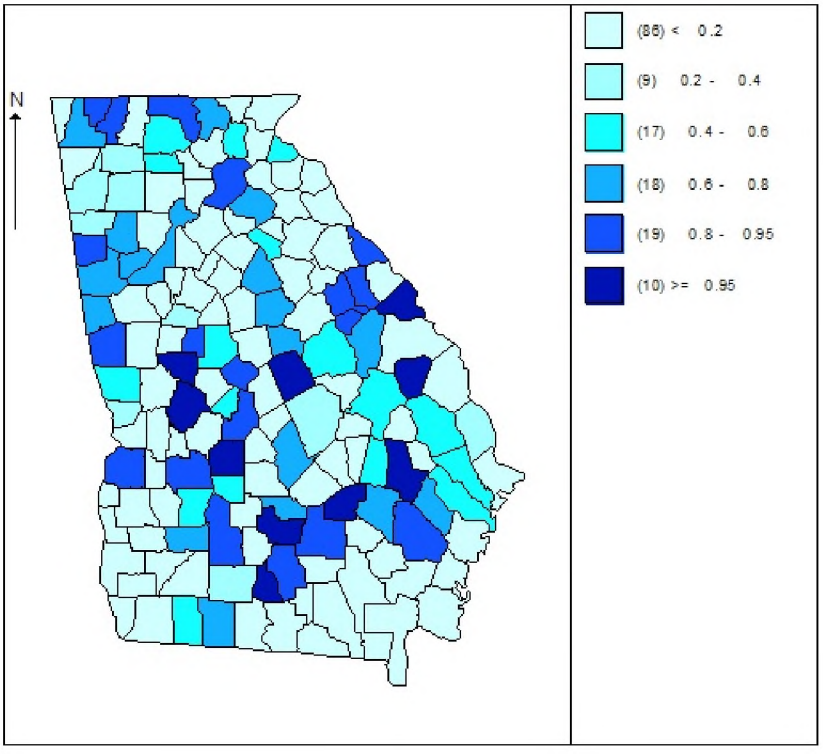


Figure 3.10: The posterior expected exceedance probability for model 3.

The disease map of the posterior expected exceedance probability for model 3 is different to the disease maps of models 1 and 2. One would expect this result as the PP follows from the relative risk values. Since model 3 has high relative risk values, the PP is also relatively high in those counties. There are 10 counties which have a PP value greater than or equal to 0.95 compared with 1 in both models 1 and 2. The high relative risks in model 3 give a false indication of the PP values for each county. One would conclude that there are a number of counties which show evidence of possible clustering when in fact this is not the case.

Table 3.16: The selected results of the posterior expected exceedance probability of model 3.

Posterior Expected Exceedance Probability			
County	**Mean**	**Standard Deviation**	**MC Error**
Richmond	0.9695	0.1719	0.001373
Jenkins	0.9800	0.1400	6.71E-4
Tattnall	0.9785	0.1451	7.59E-4
Jeff Davis	0.9956	0.0663	3.83E-4
Irwin	0.9747	0.1569	7.78E-4
Berrien	0.9345	0.2474	0.001428
Cook	0.9911	0.0940	4.99E-4
Dooly	0.9650	0.1838	9.79E-4
Taylor	0.9985	0.0384	1.93E-4
Upson	0.9654	0.1826	8.94E-4
Whitfield	0.9212	0.2694	0.001443
Wilkinson	0.9979	0.0461	2.16E-4

The selected results for the PP of model 3 show mostly the same counties with high PP values as models 1 and 2 showed. The PP values for the selected counties are significantly higher than those in both models 1 and 2 which is a result of the higher relative risk values in model 3.

3.4 South African Data

The data is collected via the civil registration system which is part of the Department of Home Affairs. Data collection is important to analyse the current health system within the country. The collection of death data is required by the Births and Deaths Registration Act introduced in 1992, (South African Government, 1992). This Act stipulates that deaths should be registered no later than 72 hours after the death took place according to Statistics South Africa (2015). These death notifications, found in Figure A.1, are collected by the Department of Home Affairs and are processed fortnightly by Statistics South Africa which produces statistical releases based on the data collected.

The South African data for acute pericarditis (Classification number: *I30*) is given in Table 3.17 along with the expected values, which are calculated based upon the province-wide incidence rate. The

disease map for the observed values is given in Figure 3.11.

Table 3.17: The observed number of deaths and expected values of acute pericarditis (Classification number: *I30*) in South Africa by province in 2014.

Province	Number of Observed Deaths	Number of Expected Deaths
Eastern Cape	3	2.0108547
Free State	1	0.8256759
Gauteng	2	3.8264741
KwaZulu-Natal	1	3.1686066
Limpopo	2	1.6682253
Mpumalanga	0	1.2530874
North-West	3	1.0892270
Northern Cape	1	0.3456705
Western Cape	3	1.8121786
Total	**16**	**16**

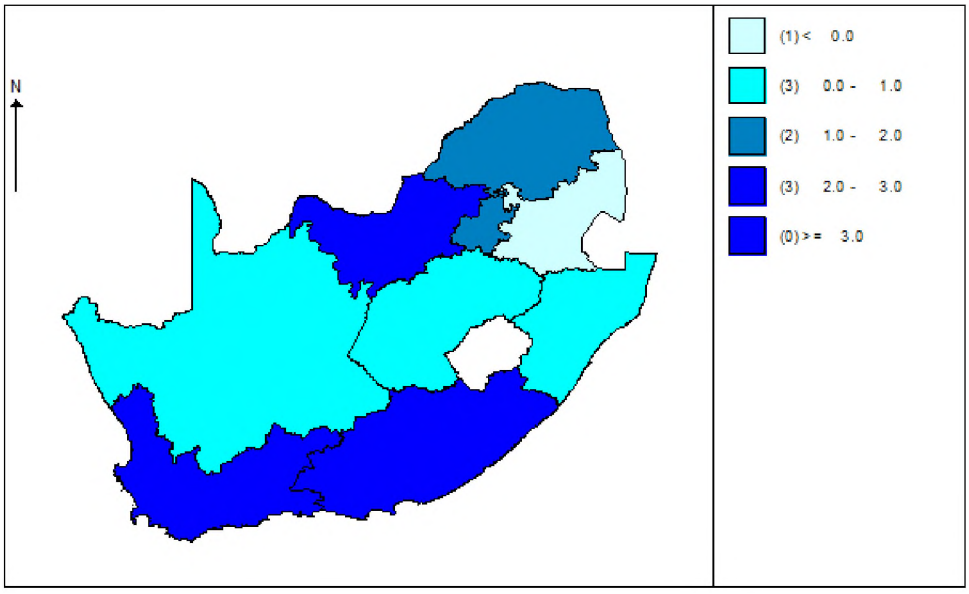


Figure 3.11: The acute pericarditis deaths in South Africa by province in 2014.

3.4.1 Acute Pericarditis

Imazio et al. (2007) state that pericarditis is caused by an inflammation of the pericardium in the heart. The pericardium is a sac which surrounds the heart and contains many blood vessels and is made up of visceral and parietal layers as stated by Lange & Hillis (2004). These layers are distanced by the pericardial cavity and contain between 15 to 50 ml of fluid. Pericarditis may cause no pain at all and may be lethargic which is common in patients who have tuberculosis, according to Lange &

Hillis (2004), but in other cases, pain may be severe and occur suddenly without warning. Pain often resonates behind the breast bone and changes depending on the position of the patient. Patients may experience relief from pain by shifting location or sitting upward. Pain may often spread to different parts of the body including the arms, neck and shoulders.

Diagnosis of acute pericarditis is challenging and is thought by Imazio et al. (2007) to be a process of exclusion.

3.4.1.1 Causes

Lange & Hillis (2004) state that in 90% of cases, the cause of acute pericarditis is unknown or thought to be a viral disease. In the other 10% of cases, the cause of acute pericarditis is believed to be "transmural myocardial infarction" along with other infections and associated with a "dissecting aortic aneurysm". Table 3.18 from Lange & Hillis (2004) shows the causes, treatments and estimated incidence for acute pericarditis.

3.4.2 Results for ICAR model

The ICAR model outlined in Section 2.12.3.1 is run for the South African data in Table 3.17. The UH component is modelled with a zero mean Gaussian distribution with a hyperprior distribution for the precision as a gamma distribution with parameters $\alpha = 0.5$ and $\beta = 0.0005$, respectively, which is the Jeffreys' prior. The precision of the ICAR prior distribution is modelled using a gamma hyperprior distribution with parameters $\alpha = 0.5$ and $\beta = 0.0005$, respectively. These parameters, $\alpha = 0.5$ and $\beta = 0.0005$, are proposed by Kelsall & Wakefield (1999) and are the parameters of the Jeffreys' prior. The model is run for $n_T = 50000$ iterations and a further $n_A = 20000$ for the calculation of the DIC. The convergence and burn-in period of the model are accessed by the BGR diagnostic and trace plots. Figure 3.12 shows the BGR plots for the deviance and the relative risk of all the provinces. The BGR is a tool to check for convergence and used to get an idea of the burn-in period, as outlined in Section 2.11.3.2.

There are four chains run for the deviance and relative risk in all provinces. These chains are run for $n_T = 50000$ iterations using different initial values. The convergence of the model is assessed by trace plots, which are given in Figures 3.13 and 3.14, and kernel densities of the relative risks and deviance. The model fit is assessed by a Q-Q plot of the deviance residuals, a histogram of the Bayesian residuals, a box plot of the ranked standardised Bayesian residuals, a box plot of the ranked deviance residuals and a plot of the Bayesian residuals versus the fitted values of the model. The deviance in both of the South African models is the saturated deviance which is calculated as the sum of the deviance of each province.

Table 3.18: The causes, treatments and estimated incidence rates for acute pericarditis.

Condition	Estimated Incidence	Clinical Indications and Tests	Treatment
Unknown	85-90%		Aspirin, NSAIDs
Infectious			
Viral	1-2%	Acute and convalescent viral titers, viral cultures, serologic test for HIV	Aspirin, NSAIDs
Bacterial	1-2%	Fever, elevated white-cell count; examination of pericardial fluid	Antibiotics, drainage of pericardial fluid
Tuberculous	4%	Chest radiography, tuberculin skin test, histologic examination, cultures, and measurement of adenosine deaminase level in pericardial fluid and tissue	Multidrug antituberculous therapy and prednisone
Acute Myocardial Infarction	NA	Electrocardiography, serum troponin or creatine kinase, echocardiogram	Aspirin (avoid NSAIDs)
Aortic Dissection	<1%	Magnetic resonance imaging, computed tomography, transesophageal echocardiography	Urgent surgery
Trauma	NA	Clinical history	NSAIDs (avoid aspirin)
Neoplasm	7%	Constitutional symptoms, lymphadenopathy; chest radiography, examination of pericardial fluid	NSAIDs, glucocorticoids (by intrapericardial instillation)
Chest-wall Irradiation	<1%	Clinical history	NSAIDs
Uremia	NA	Serum blood urea nitrogen and creatinine levels	Initiate or intensify dialysis
Cardiotomy or Thoracic Surgery	<1%	Clinical history, evidence of polyserositis; chest radiography, erythrocyte sedimentation rate	Aspirin, NSAIDs
Autoimmune or Inflammatory Disease	3-5%	Rheumatoid factor, complement levels, antinuclear antibodies	Aspirin, NSAIDs, glucocorticoids
Adverse Drug Reaction	<1%	Clinical history; eosinophil count	Discontinue drug; aspirin, NSAIDs

Where NSAID is a nonsteroidal anti-inflammatory drug.

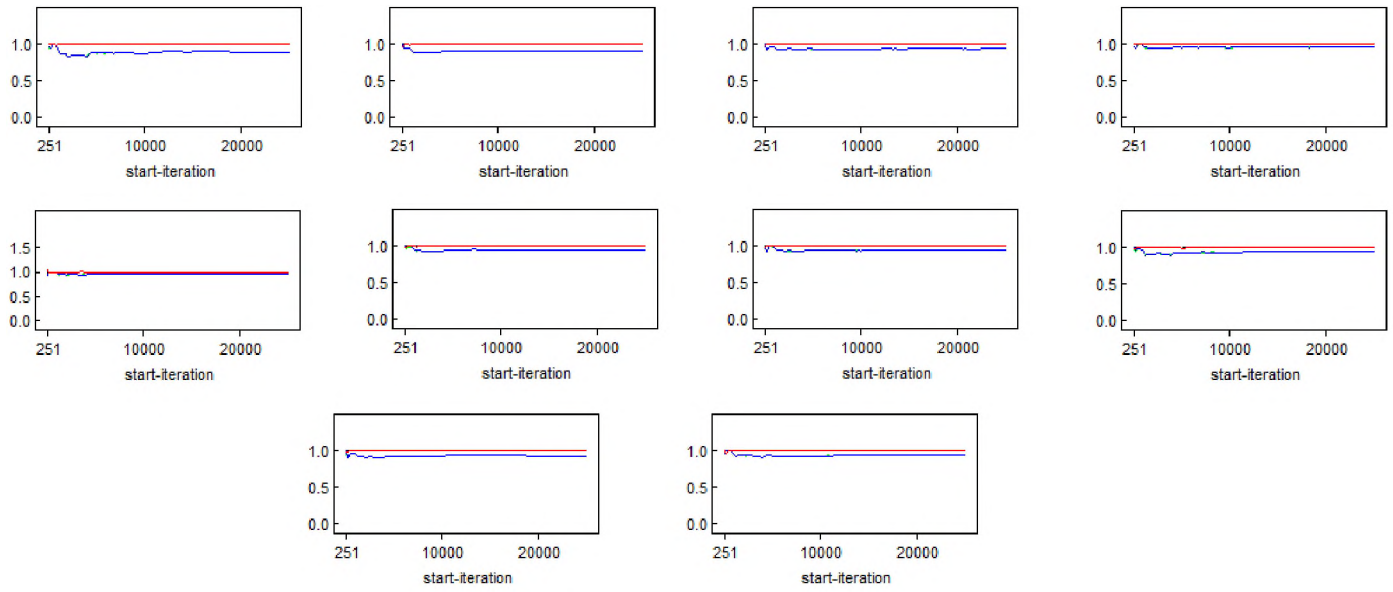


Figure 3.12: The BGR plots for the ICAR model from top left: Deviance, Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape and Western Cape.

Figure 3.12 shows that the model does converge as the red line, which is the ratio of pooled and within, appears to be stable around the value of 1. The blue line also appears to be stable after $n_T = 10000$ iterations.

Table 3.19: The statistics from the BGR for the ICAR model.

Iteration Range	Unnormalised		Normalised as plotted		BGR ratio
	of pooled chains	mean within chain	of pooled chains	mean within chain	
251–500	5.217	5.316	0.9814	1.0000	0.9814
501–1000	3.977	3.967	0.7480	0.7461	1.0030
751–1500	3.881	3.846	0.7300	0.7233	1.0090
1001–2000	3.713	3.743	0.6985	0.7040	0.9921
1251–2500	3.976	4.000	0.7478	0.7524	0.9939
1501–3000	4.277	4.309	0.8046	0.8105	0.9927
1751–3500	4.097	4.158	0.7706	0.7821	0.9853
2001–4000	4.118	4.150	0.7745	0.7807	0.9921
2251–4500	4.090	4.076	0.7693	0.7667	1.0030
2501–5000	4.050	4.050	0.7617	0.7618	0.9999
2751–5500	3.838	3.834	0.7220	0.7212	1.0010
3001–6000	3.792	3.791	0.7132	0.7130	1.0000
3251–6500	3.888	3.885	0.7313	0.7308	1.0010
3501–7000	3.908	3.903	0.7351	0.7342	1.0010
3751–7500	3.864	3.860	0.7267	0.7261	1.0010
4001–8000	3.902	3.920	0.7340	0.7374	0.9953
4251–8500	3.918	3.904	0.7369	0.7342	1.0040
4501–9000	3.844	3.832	0.7230	0.7209	1.0030
4751–9500	3.852	3.826	0.7244	0.7197	1.0070
5001–10000	3.833	3.814	0.7210	0.7174	1.0050
5251–10500	3.862	3.843	0.7265	0.7229	1.0050
5501–11000	3.916	3.905	0.7366	0.7346	1.0030
5751–11500	4.000	3.996	0.7524	0.7517	1.0010
6001–12000	4.012	4.012	0.7547	0.7546	1.0000

Table 3.19 shows the results of the BGR convergence diagnostic applied to the deviance and relative risks of the model. From the BGR ratio, it may be seen that the model reaches convergence between 501-1000 iterations as this is the first time that the BGR ratio reaches 1. Gelman & Hill (2007) propose that a BGR ratio of 1.1 is an "acceptable limit" for convergence. The burn-in period may also be determined from Table 3.19. Although interpretation may be slightly subjective, the burn-in period is between iterations 5751-11500 as the statistics in the table become more stable around these iterations. The burn-in period is therefore set at $N_B = 12000$ as this is above 11500 and ensures stability in the statistics.

Table 3.20: An analysis of the burn-in period for the ICAR model.

Province	Percentage
Eastern Cape	0.7517
Free State	0.5317
Gauteng	0.7660
KwaZulu-Natal	1.0253
Limpopo	0.5530
Mpumalanga	0.9240
North-West	1.2775
Northern Cape	1.0941
Western Cape	1.2102

The proportion of the MC error to the standard deviation of the relative risks of the ICAR model are used to assess whether the burn-in period of $N_B = 12000$ is sufficient. This proportion, expressed as a percentage in Table 3.20, has been calculated for all the relative risks of the provinces and all of the percentages are below 5% which indicates that the burn-in period of $N_B = 12000$ is sufficient.

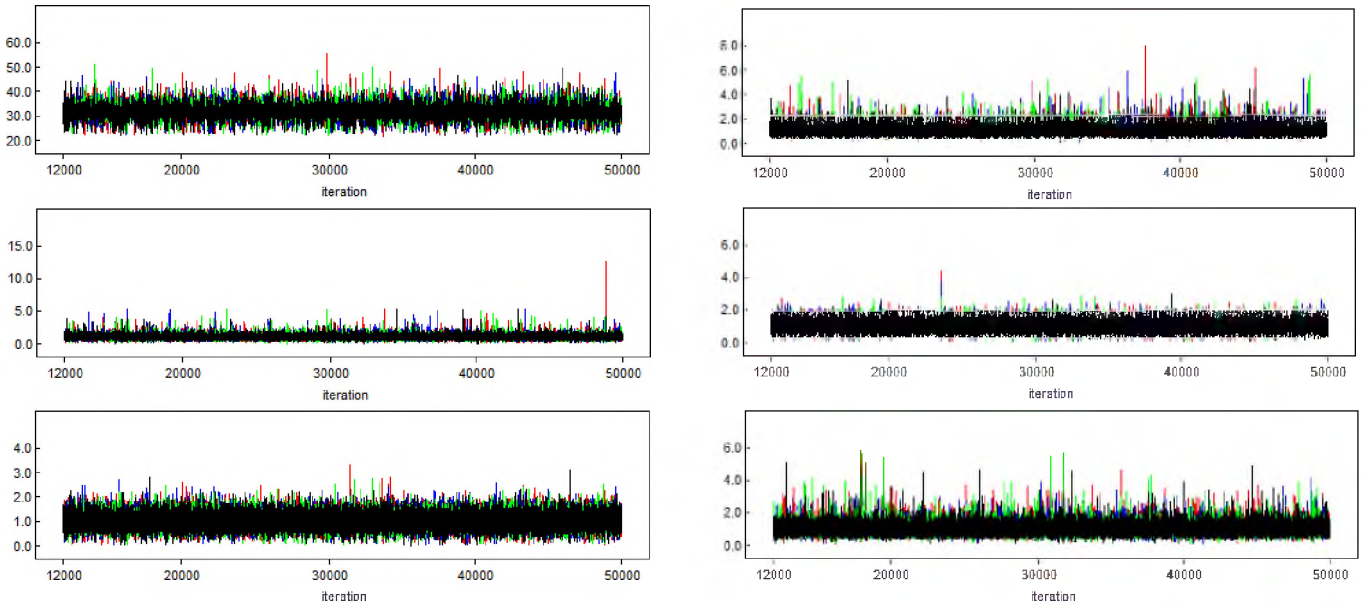


Figure 3.13: The trace plots for multiple chains for the South African data for the ICAR model from top left: Deviance, Eastern Cape, Free State, Gauteng, KwaZulu-Natal and Limpopo.

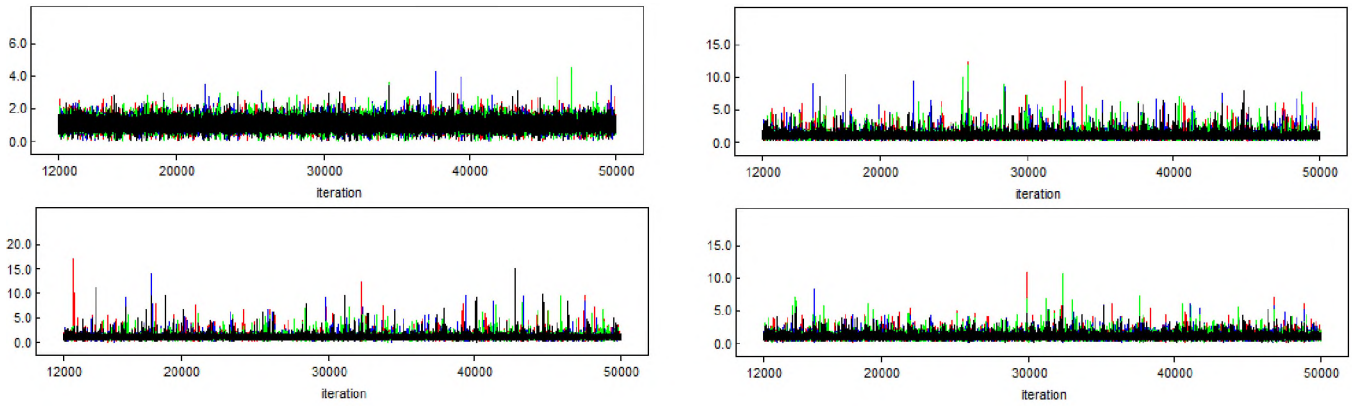


Figure 3.14: The trace plots for multiple chains for the South African data for the ICAR model continued from top left: Mpumalanga, North-West, Northern Cape and Western Cape.

Figures 3.13 and 3.14 show the trace plots for the four chains for the relative risk of all the provinces and the deviance of the ICAR model. The trace plots show that convergence is reached in the relative risks of all the provinces as well as the deviance. This may be shown by covering the trace plots by a "thick pen." All four chains remain stable within an interval.

The model was run with a single chain for $n_T = 50000$ iterations with a burn-in period of $N_B = 12000$ iterations and the following results were produced:

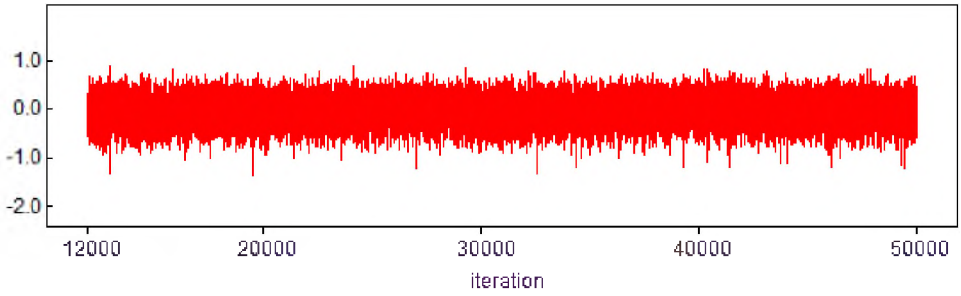


Figure 3.15: The trace plot for the intercept term in the ICAR model.

The intercept term β_0 converges as is shown in the trace plot. The intercept term is needed as the random effects in the model sum to 0.

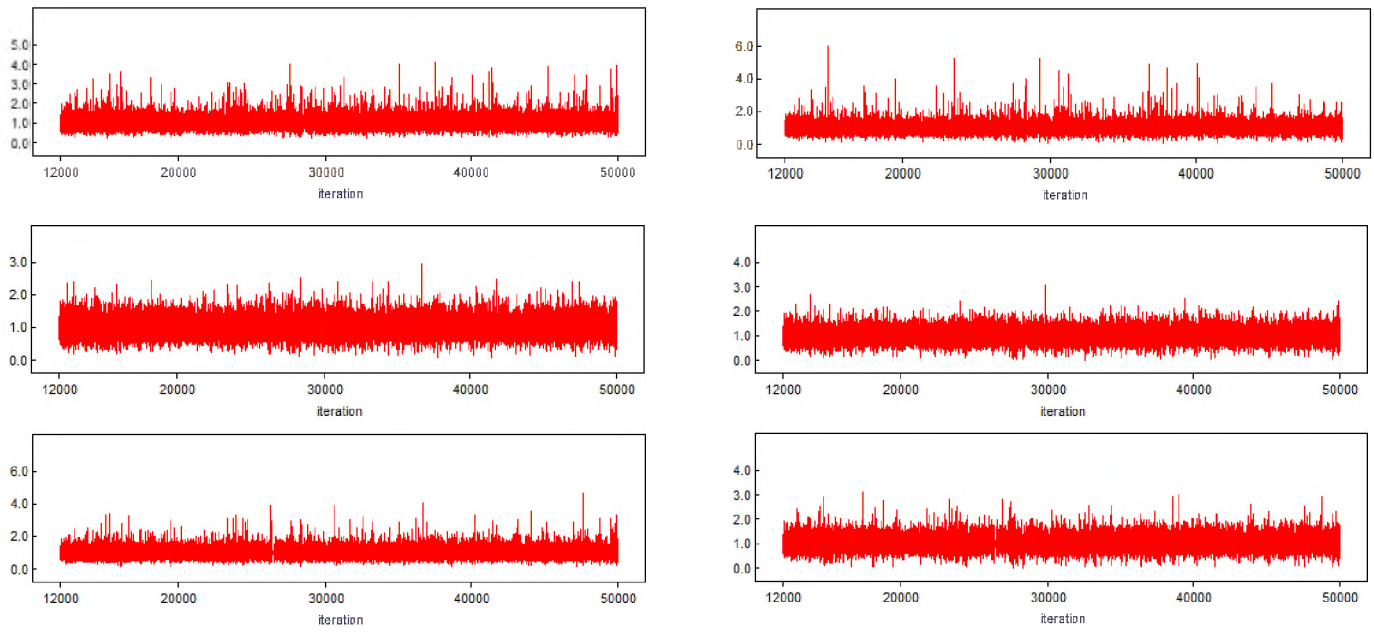


Figure 3.16: The trace plots for South African data for the ICAR model from top left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo and Mpumalanga.

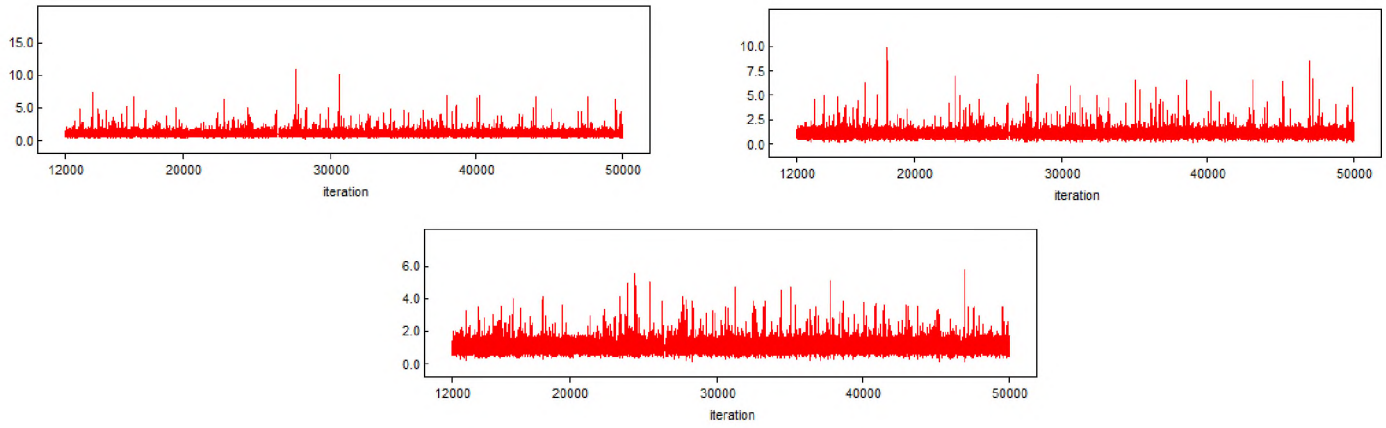


Figure 3.17: The trace plots for South African data for the ICAR model continued from top left: North-West, Northern Cape and Western Cape.

Figures 3.16 and 3.17 also show the convergence of the relative risks of the ICAR model which is to be expected as convergence was shown for multiple chains.

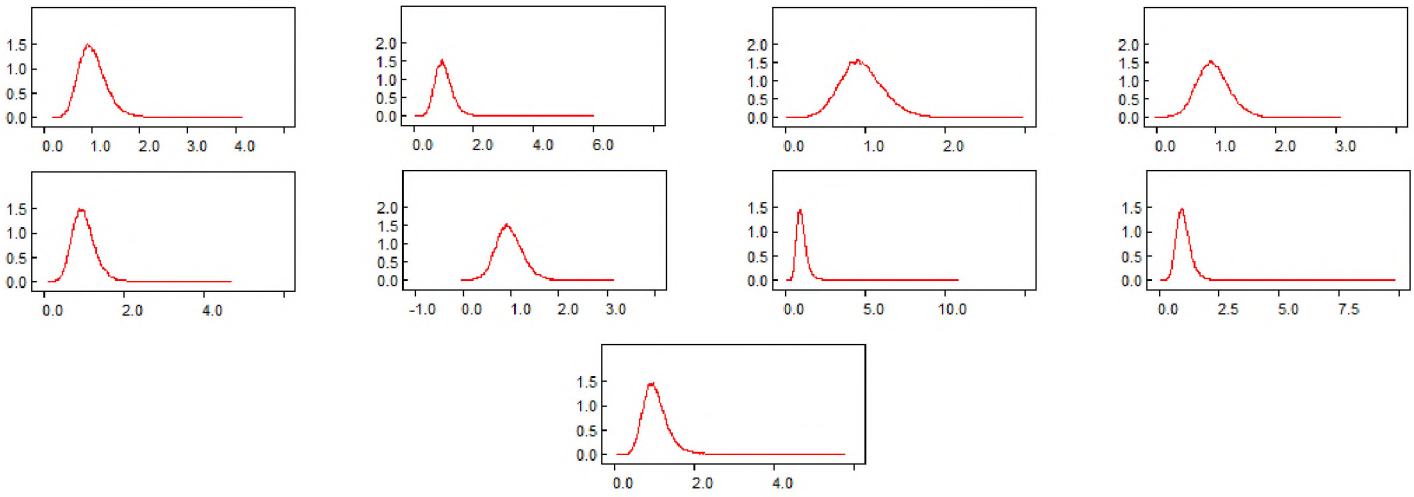


Figure 3.18: The density kernel for the relative risks of the South African data for the ICAR model from top left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga North-West, Northern Cape and Western Cape.

The density kernels for the relative risks are shown in Figure 3.18. The density kernel plot gives an estimate of the shape of the marginal posterior distribution for each relative risk. The densities have a heavy tail. All of the densities seem to have a maximum around 1.

Table 3.21: The model diagnostics of the South African data for the ICAR model.

Statistic	Mean	Standard Deviation	MC Error	95% Confidence Interval	
				2.5%	97.5%
MAPE	1.461	0.3972	0.002155	0.7778	2.333
MSPE	3.774	2.343	0.01192	1.111	9.889
pE	0.9569	0.09267	0.001765	0.6207	0.9993
pV	0.02647	0.07265	0.001765	2.416E-4	0.2711
QR60	1.321	0.7269	0.0145	1.031	3.063
pS	0.01667	0.06089	0.001737	6.883E-5	0.2245
Saturated Deviance	10.44	1.928	0.02203	6.451	14.95
DIC	32.074				
pD	1.45				
Dhat	29.174				
Dbar	30.624				

The MAPE and MSPE of the ICAR model are relatively low values and are similar to those calculated for the PCAR model.

The quantile ratio (QR) is calculated for the relative risks and is used to assess the between-province variation in the relative risks. The QR shows the extent of the spread of the "empirical distribution" of the relative risks as in Lunn et al. (2012). The 60% QR, denoted by QR60, is calculated for the model by using the $ranked()$ function in WinBUGS®, which will rank the relative risks of the provinces

and then the calculation involves estimating the "exponentiated difference" between the relative risks at the provinces with 80% and 20% quantiles which are then ranked. The code to calculate QR60 is: $QR60 < -ranked(rr[], 8)/ranked(rr[], 2)$.

The QR is calculated as 1.321 which indicates that there is a relatively low heterogeneity in the relative risks of the middle 60% of the provinces in the model.

The amount of spatial to unstructured variation in the model is assessed. The variation contributed by the unobserved spatial factors is 1.667% of the total variation in the model, this is denoted by pS in Table 3.21. The variation contributed by the unobserved non-spatial factors, denoted by pV, is 2.647% of the total variation, with the other 95.69%, pE, of the total variation is due to the expected values of the provinces.

The DIC is 32.074, resulting from the model being run for an additional $n_A = 20000$ iterations. The posterior mean of the deviance, $Dbar$, is 29.174.

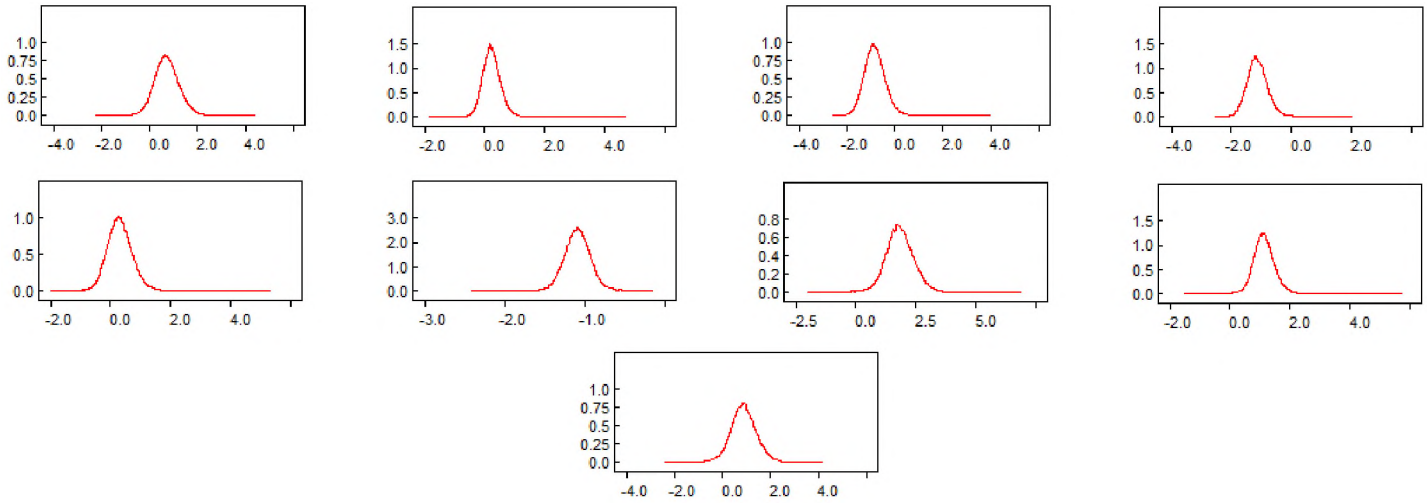


Figure 3.19: The density kernels for the standardised Bayesian residuals of the South African data ICAR model from top left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape and Western Cape.

Lawson (2009) proposes two features of residuals in a model when the model fits the data well. The first feature is that the residuals should be symmetric and centred around 0. Although the centring around 0 is an approximation due to the nature of the simulation in the model. The second feature is that the residuals must show a random pattern and show no particular structure. Figure 3.19 shows the density kernels for the residuals in each province in the South African ICAR model. Based on the Figure, Eastern Cape, Free State and Limpopo appear to be approximately centred around 0. KwaZulu-Natal, Mpumalanga and North-West have centres which are further away from 0. An approximate symmetry of the kernel densities is shown in Eastern Cape, Free State, North-West and Western Cape which may indicate a good model fit for these provinces. The symmetry of the other provinces appears to be skewed.

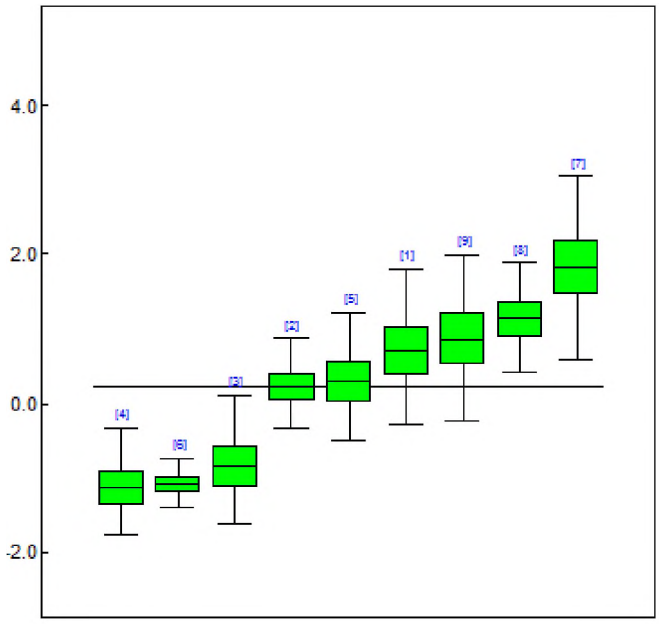


Figure 3.20: The ranked box plots of the standardised Bayesian residuals of the South African data for the ICAR model from left to right: KwaZulu-Natal, Mpumalanga, Gauteng, Free State, Limpopo, Eastern Cape, Western Cape, Northern Cape and North-West.

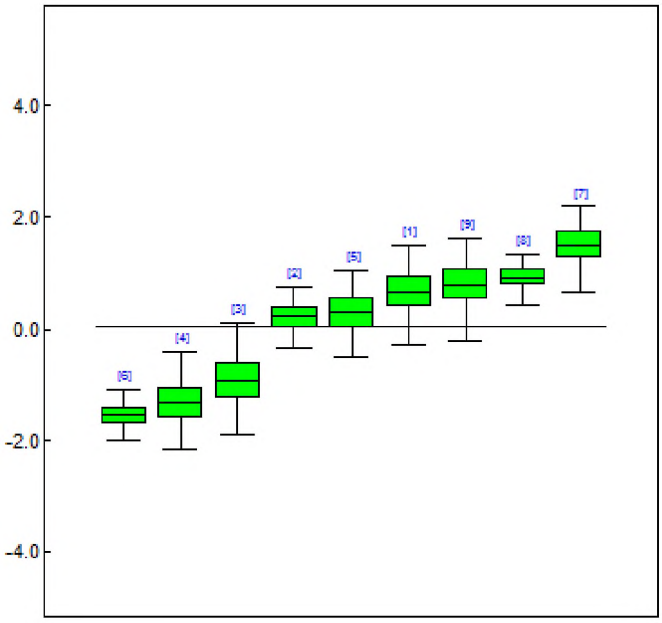


Figure 3.21: The ranked box plots of the deviance residuals of the South African data for the ICAR model from left to right: Mpumalanga, KwaZulu-Natal, Gauteng, Free State, Limpopo, Eastern Cape, Western Cape, Northern Cape and North-West.

Figure 3.20 shows the ranked box plots of the standardised Bayesian residuals in the South African

ICAR model. The ranking is achieved in WinBUGS® by plotting the box plot and then ranking the box plots by right-clicking on the plot and selecting $properties$ and then selecting $special$ and selecting the box named $ranked$. WinBUGS® provides two options for the ranking of the box plots. The first option is to rank by the mean and the second option is to rank by the median. Figures 3.20 and 3.21 makes use of the ranking by mean option. This is the method proposed by Lunn et al. (2012).

We expect all of the standardised Bayesian residuals to occur between, as explained in Section 2.13.4, -2 and 2. The box plots of all the provinces in Figures 3.20 and 3.21 show that all of the standardised Bayesian residuals do occur inside this bound. North-West is very close to occurring outside this bound. The standardised Bayesian residuals that occur outside this bound are then classified as outliers. Since none of the box plots of the standardised Bayesian residuals in the ICAR model occur outside of this bound, there are no Bayesian residuals which can be classified as an outlier.

The ranked box plots of the deviance residuals also show that all of the deviance residuals are between -2 and 2. The whiskers of the box plot of the deviance residuals for the North-West are much shorter than the box plot of North-West for the standardised Bayesian residuals.

Table 3.22: The statistics from the Bayesian residuals of the South African data for the ICAR model.

Province	**Bayesian Residual**			**95% Confidence Interval**	
	Mean	**Standard Deviation**	**MC error**	**2.5%**	**97.5%**
Eastern Cape	0.9258	0.6194	0.004785	-0.4678	1.8980
Free State	0.1650	0.2568	0.001421	-0.3893	0.5735
Gauteng	-1.7100	1.0520	0.007847	-3.9750	0.1699
KwaZulu-Natal	-2.0370	0.8872	0.008782	-3.9250	-0.4344
Limpopo	0.3113	0.5091	0.002780	-0.8099	1.1320
Mpumalanga	-1.2150	0.3600	0.003093	-1.9900	-0.5715
North-West	1.8380	0.4216	0.005140	0.9020	2.3880
Northern Cape	0.6370	0.1316	0.001392	0.3584	0.8140
Western Cape	1.0820	0.6464	0.008066	-0.4070	2.0000

Table 3.23: The deviance residuals of the ICAR model.

Province	Deviance Residual			95% Confidence Interval	
	Mean	Standard Deviation	MC error	2.5%	97.5%
Eastern Cape	0.6448	0.4425	0.002901	-0.2644	1.477
Free State	0.2039	0.2787	0.001473	-0.3455	0.740
Gauteng	-0.9309	0.5047	0.004340	-1.8940	0.096
KwaZulu-Natal	-1.3240	0.4469	0.004640	-2.1610	-0.397
Limpopo	0.2725	0.3928	0.002003	-0.5093	1.033
Mpumalanga	-1.5430	0.2334	0.002103	-2.0010	-1.074
North-West	1.4730	0.4009	0.003986	0.6212	2.184
Northern Cape	0.8958	0.2338	0.001869	0.4195	1.315
Western Cape	0.7723	0.4586	0.004218	-0.2000	1.606

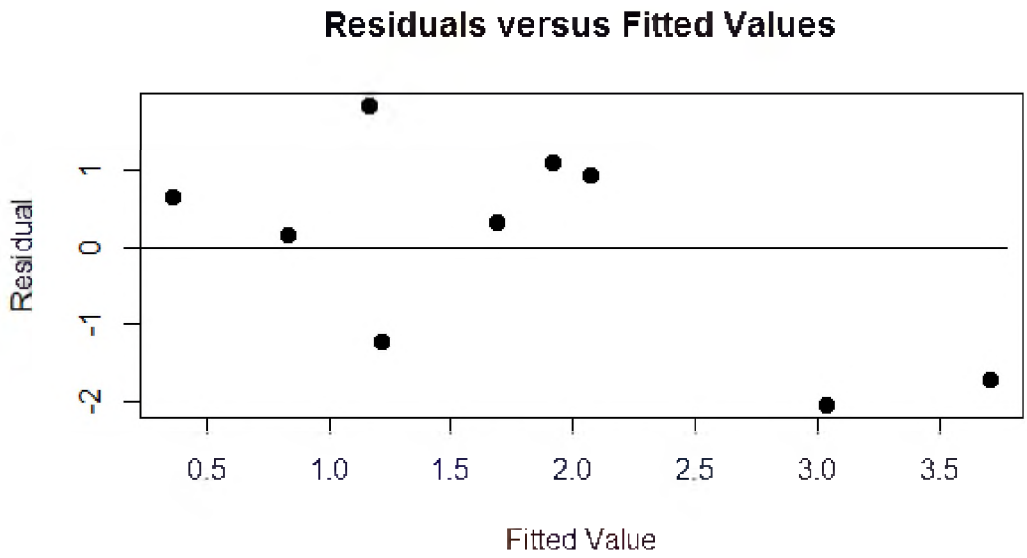


Figure 3.22: The Bayesian residuals versus the fitted values of the ICAR model.

The second feature shown by residuals when a model fits the data well, proposed by Lawson (2009), is that there should be a random pattern in the residuals. Figure 3.22 shows the plot of the Bayesian residuals versus the fitted values for the ICAR model. Based on the figure, the pattern shown by the Bayesian residuals is reasonably random although the pattern of the Bayesian residuals after 2.5 shows an increase in the space between the Bayesian residuals. These two provinces, after a fitted value of 2.5, are Gauteng and KwaZulu-Natal which show that the model may not fit well to these provinces. KwaZulu-Natal seems to have a poor model fit and this is shown throughout the plots.

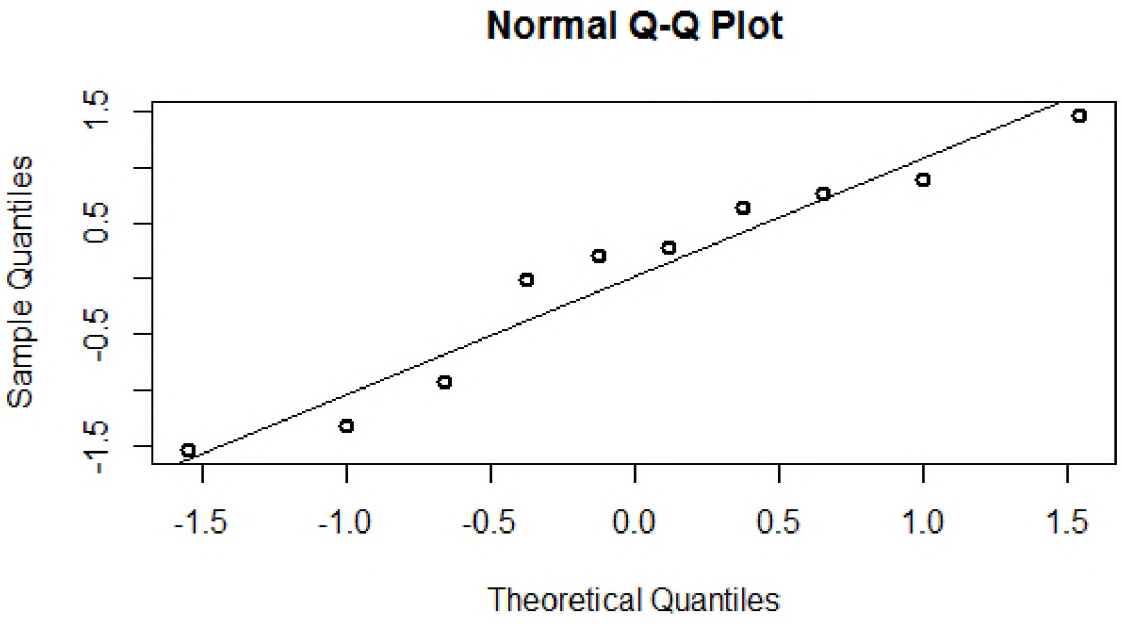


Figure 3.23: Q-Q plot of the deviance residuals of the ICAR model.

The Q-Q plot of the deviance residuals for the ICAR model shows a reasonable fit for most of the provinces. Most of the provinces show adequate model fit although some provinces are further away from the straight line than other provinces are. Based on the two features proposed by Lawson (2009), the model fits the data adequately.

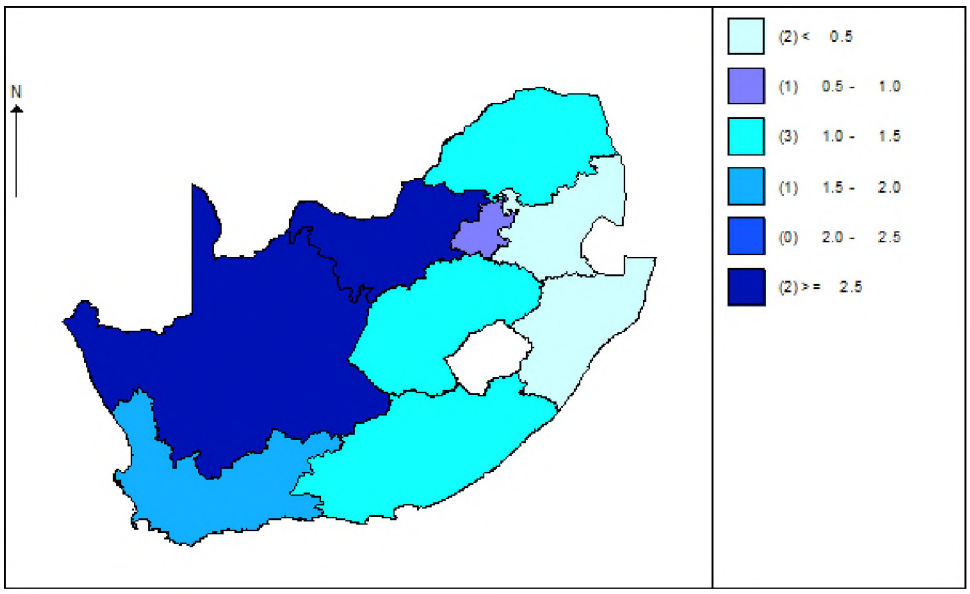


Figure 3.24: The SMR of the South African data for the ICAR model.

KwaZulu-Natal and Mpumalanga have the lowest SMR values out of all the provinces. There are

1 and 0 observed deaths with expected number of deaths of 3.1686 and 1.253 for KwaZulu-Natal and Mpumalanga, respectively. A low number of observed deaths and relatively high expected values contribute to the low values for SMR. The high expected value for KwaZulu-Natal results from the large population of the province at 10694434, which is the second highest population of a province in South Africa after Gauteng. The combination of a large population size and a small number of observed deaths result in a small SMR value. Mpumalanga has the sixth largest population size of all the provinces in South Africa but since the observed number of deaths is 0, the SMR is also approximately 0. The highest SMR values are at the North-west and Northern Cape with 3 and 1 observed deaths and expected number of deaths of 1.0892 and 0.3457 respectively. The low expected number of deaths with respect to the number of observed deaths result in a high SMR value. The low expected number of deaths is a result of a small population size in these provinces. The Northern Cape and North-West have population sizes of 1166680 and 3676274, respectively, which are the smallest and third smallest populations, respectively. Therefore the relative risk is a better estimate of the risk of a disease than the SMR is, as the relative risk is not a function of the expected values whereas SMR is calculated based on the expected values and thus is influenced by the population size of the province.

Table 3.24: The SMR of the South African data for the ICAR model.

Province	SMR			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Eastern Cape	1.4920	0.007775	0.00004145	1.477	1.508
Free State	1.2110	0.008173	0.00004347	1.196	1.228
Gauteng	0.5226	0.003961	0.00002114	0.515	0.530
KwaZulu-Natal	0.3155	0.006860	0.00003661	0.302	0.329
Limpopo	1.1990	0.003791	0.00002021	1.192	1.207
Mpumalanga	0.0008	0.025410	0.0001354	-0.052	0.047
North-West	2.7560	0.051330	0.0002733	2.660	2.861
Northern Cape	2.910	0.180500	0.0009529	2.606	3.309
Western Cape	1.6560	0.011500	0.00006131	1.634	1.679

Table 3.24 shows the results from the MCMC simulation with $n_T = 50000$ iterations. The SMR for Mpumalanga is 0.0008 which is expected as the observed number of deaths is 0. The SMR would usually be 0 in this province as the observed value is 0 but because the SMR has a prior distribution, this is not the case although the approximation is very close. The Northern Cape, which has the lowest expected number of deaths in the model, has the highest SMR value which is attributed to the low expected number of deaths.

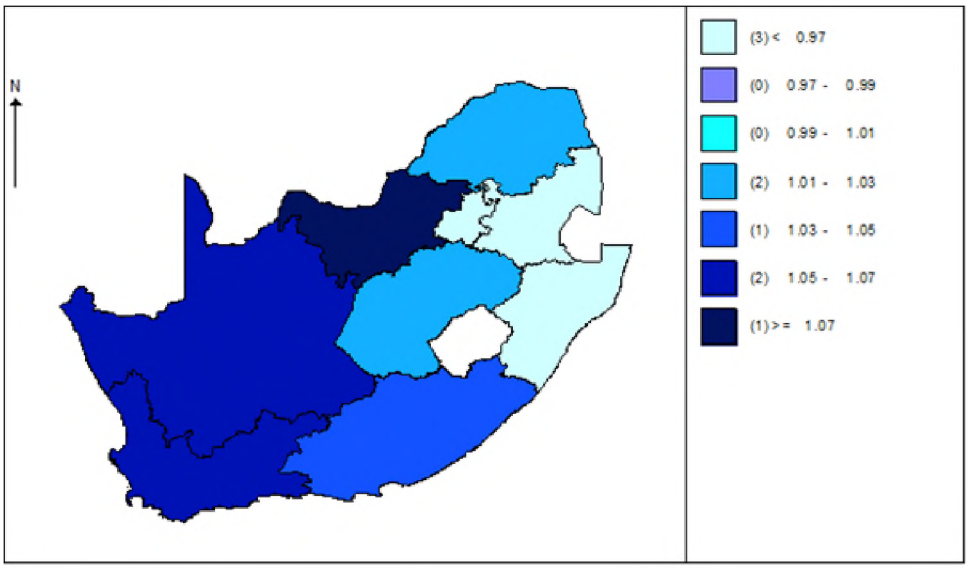


Figure 3.25: The relative risk of the South African data for the ICAR model.

Table 3.25: The relative risk of the South African data for the ICAR model.

Province	Relative Risk			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Eastern Cape	1.0320	0.3113	0.002340	0.5526	1.7410
Free State	1.0110	0.3138	0.001669	0.5164	1.6990
Gauteng	0.9677	0.2756	0.002111	0.4833	1.5650
KwaZulu-Natal	0.9547	0.2808	0.002879	0.4382	1.5590
Limpopo	1.0110	0.3065	0.001695	0.5183	1.7010
Mpumalanga	0.9666	0.2894	0.002674	0.4478	1.5810
North-West	1.0710	0.3997	0.005106	0.5587	1.9510
Northern Cape	1.0520	0.3901	0.004268	0.5375	1.8940
Western Cape	1.0590	0.3582	0.004335	0.5547	1.8780

The relative risk of acquiring and dying from acute pericarditis is shown in Figure 3.25 and Table 3.25. KwaZulu-Natal, Mpumalanga and Gauteng have the lowest relative risk of all the provinces and are all below 0.97. This result is different from the SMR values as Mpumalanga had the lowest SMR value but does not have the lowest relative risk value which is now KwaZulu-Natal. Although KwaZulu-Natal had the second smallest SMR value. Western Cape, Northern Cape and North-West have high relative risk values with the highest associated with North-West at 1.071. The SMR was highest at the Northern Cape which is now third highest for the relative risk. All of the relative risks are not far away from 1 which is regarded as the "null risk value" and resulted in a low PP. It can thus be said that the risk of contracting and dying from acute pericarditis in South Africa is low when using the ICAR model.

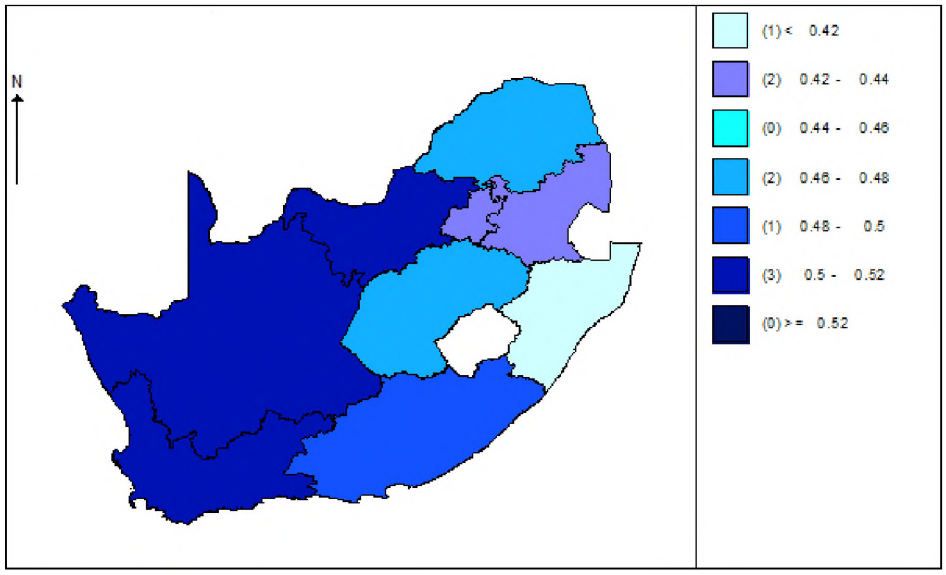


Figure 3.26: The posterior exceedance probability of South African data for the ICAR model.

Table 3.26: The posterior exceedance probability of the South African data for the ICAR model.

Posterior Expected Exceedance Probability			
Province	**Mean**	**Standard Deviation**	**MC Error**
Eastern Cape	0.4925	0.4999	0.002652
Free State	0.4690	0.4990	0.002522
Gauteng	0.4235	0.4941	0.003162
KwaZulu-Natal	0.4098	0.4918	0.003744
Limpopo	0.4701	0.4991	0.002553
Mpumalanga	0.4235	0.4941	0.003305
North-West	0.5131	0.4998	0.003337
Northern Cape	0.5025	0.5000	0.003108
Western Cape	0.5084	0.4999	0.003646

Using the notation from Section 2.13.3, let the threshold level, denoted by c, be equal to 1. This is a fair number as it is the "null risk value." The PP is then the posterior expected exceedance probability calculated as the proportion of relative risk values which surpass the threshold of 1. Figure 3.26 shows the PP for all provinces in the model and it can be seen that the PP is below 0.52 for all provinces. This means that the province with the highest relative risk value only surpasses 1 less than 52% of the time. The highest values of the PP occur at Western Cape, Northern Cape and North-west. This is expected as these 3 provinces have the highest SMR and relative risk values.

If the exceedance probability of interest, $\hat{P}$, is as high as 0.9 then Figure 3.26 shows that there are no unusual or unexpected clusters of the disease in the model. There will also be no provinces with a higher than expected relative risk. Lowering the exceedance probability of interest, $\hat{P}$, to 0.5 will

result in unusual or unexpected clusters of the disease in the model but is not commonly done as this information is not useful as $\hat{P}$ is low.

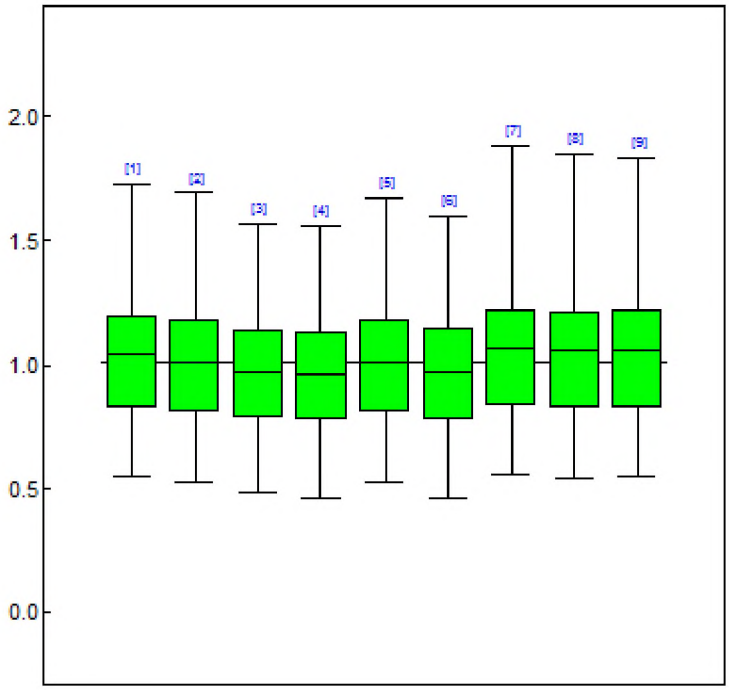


Figure 3.27: The box plot of the relative risk of the South African data for the ICAR model from the left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape and Western Cape.

The box plot of the relative risks shows that the relative risks are mostly centred around 1 with North-West, Western Cape and Northern Cape just above 1. The last three box plots which correspond to North-West, Northern Cape and Western Cape do show a long tail above the value of 1, which is expected as the standard error for these 3 provinces is higher than the standard deviations of all the other provinces.

Additional figures are available in Appendix C.4.

3.4.3 Results for PCAR model

The PCAR model is given in Section 2.12.3.2 and is also based on the data in Table 3.17. The precision for the PCAR prior distribution is assigned a gamma hyperprior distribution with parameters of $\alpha = 0.5$ and $\beta = 0.0005$, respectively, which are the parameters of the Jeffreys' prior. The UH component is assigned a zero mean Gaussian distribution with the hyperprior of the precision modelled using a gamma distribution with parameters of $\alpha = 0.5$ and $\beta = 0.0005$, respectively, which are again the parameters of the Jeffreys' prior. These are the same parameters as the ICAR model in Section 3.4.2.

Model assessment is performed in the same way as the ICAR model in Section 3.4.2. The BGR plots for the deviance and the relative risk in all provinces for the PCAR model are given in Figure 3.28. The model is run for $n_T = 50000$ iterations and four chains are used with four different sets of initial values, as with the ICAR model in Section 3.4.2.

The implementation of the PCAR model is based on the method by Spiegelhalter et al. (2003). The PCAR model is implemented in WinBUGS® by firstly calculating m_{ii} which are the elements proportional to a conditional covariance $S_i|S_j$ and are contained in the vector $\boldsymbol{M}$. The elements m_{ii} are calculated using Equation 2.22 which is written in WinBUGS® as: $M[i] < -1/e[i]$. The next step is to calculate the cumulative sum of the neighbours vector. This is the same cumulative sum that was calculated for the Moran's Autocorrelation coefficient in the Georgia models although this is done in WinBUGS® for the PCAR model and not in RStudio®. The first element of the cumulative sum is 0 which is implemented as: $csum[1] < -0$ and the rest of the cumulative sum is calculated by starting at the second element using a *loop* as:

$$\begin{aligned} for\ (i\ in\ 2:(N+1)) \quad & \{ \\ csum[i] \ <- \ & sum(num[1:(i-1)])\}, \end{aligned}$$

where $\boldsymbol{num}$ is the vector of the number of neighbours for each province. The next step is to determine an index matrix, denoted by $pick[]$, which will contain the index of the elements of the weight vector $\boldsymbol{W}$ and will have N columns and number of rows equal to the number of elements in $\boldsymbol{W}$. The $pick[]$ matrix is calculated such that the i^{th} column contains a value of 1 in all of the J rows such that, $csum[i] < J \leq csum[i+1]$ and 0 otherwise. This is done such that, $\boldsymbol{W}$ at the elements of $W[(csum[i]+1) : csum[i+1]]$, are the set of weights w_{ij} corresponding to the spatial correlation between provinces i and j. The code for the $pick[]$ matrix is:

$$\begin{aligned} for(k\ in\ 1:sumNumNeigh) \quad & \{ \\ for(i\ in\ 1:N) \quad & \{ \\ pick[k,i] \ <- \ & step(k - csum[i] - epsilon) * step(csum[i+1] - k), \end{aligned}$$

where $epsilon$ is equal to 0.0001.

The vector $\boldsymbol{W}$ can now be calculated using the $inprod()$ function in WinBUGS®. The $inprod()$ function is used to calculate the inner product of two vectors. The $inprod()$ function is used to determine which province is associated with which element in $\boldsymbol{W}$. This is done using the inner product of the k^{th} row of the $pick[]$ matrix and the expected value, which is the denominator in Equation 2.21,

given as:

$$w_{ij} = \begin{cases} \left(\frac{e_i}{e_j}\right)^{\frac{1}{2}} & \text{if regions } i \text{ and } j \text{ are adjacent} \\ 0 & \text{Otherwise.} \end{cases}$$

Thus the $inprod()$ function results in 0 and 1 elements which will determine which provinces are adjacent. The last part of the calculation of the vector $\boldsymbol{W}$ is to take the square root of the proportion of the expected values of the adjacent provinces. The element k of the vector $\boldsymbol{W}$ is thus calculated as:

$$W[k] < -sqrt(e[adj[k]]/inprod(e[], pick[k,])).$$

The PCAR prior can now be implemented using:

$$S[1:N] \sim car.proper(mu[], W[], adj[], num[], M[], omega, \rho),$$

where $mu[]$ is assigned a prior distribution as: $alpha \sim dnorm(0, 0.0001)$ and $omega$ is assigned a Jeffreys' prior and ρ has a uniform distribution which is bounded in WinBUGS® as:

$$\begin{aligned} rho.min &< - \quad min.bound(W[], adj[], num[], M[]) \\ rho.max &< - \quad max.bound(W[], adj[], num[], M[]). \end{aligned}$$

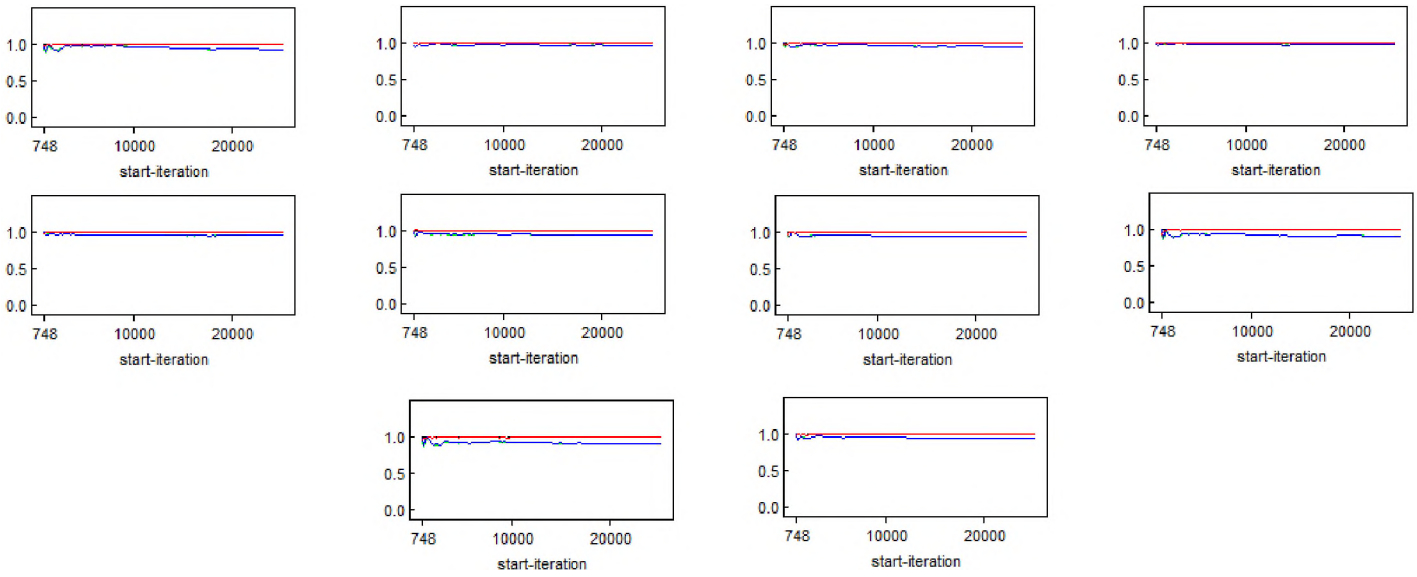


Figure 3.28: The BGR plots of the PCAR model from top left: Deviance, Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape and Western Cape.

Table 3.27: The statistics from the BGR of the PCAR model.

Iteration Range	Unnormalised		Normalised as plotted		BGR ratio
	of pooled chains	mean within chain	of pooled chains	mean within chain	
748–995	3.944	4.007	0.9579	0.9732	0.9842
996–1490	3.679	3.728	0.8935	0.9055	0.9868
1243–1985	4.028	4.046	0.9783	0.9827	0.9956
1491–2480	3.934	3.992	0.9554	0.9696	0.9854
1738–2975	3.814	3.826	0.9262	0.9292	0.9968
1986–3470	3.840	3.817	0.9326	0.9270	1.0060
2233–3965	3.794	3.768	0.9213	0.9150	1.0070
2481–4460	3.928	3.899	0.9538	0.9469	1.0070
2728–4955	3.928	3.918	0.9540	0.9515	1.0030
2976–5450	4.083	4.048	0.9915	0.9830	1.0090
3223–5945	4.063	4.028	0.9867	0.9782	1.0090
3471–6440	3.998	3.976	0.9708	0.9657	1.0050
3718–6935	4.048	4.046	0.9830	0.9825	1.0010
3966–7430	4.070	4.050	0.9883	0.9836	1.0050
4213–7925	4.033	4.038	0.9795	0.9806	0.9989
4461–8420	4.020	4.033	0.9762	0.9795	0.9966
4708–8915	4.012	4.035	0.9743	0.9799	0.9943
4956–9410	4.057	4.072	0.9852	0.9889	0.9963
5203–9905	4.063	4.069	0.9867	0.9882	0.9985
5451–10400	3.998	3.999	0.9709	0.9711	0.9998
5698–10895	4.001	3.996	0.9718	0.9703	1.0010
5946–11390	4.054	4.041	0.9844	0.9815	1.0030
6193–11885	4.063	4.060	0.9868	0.9859	1.0010
6441–12380	4.062	4.055	0.9866	0.9848	1.0020

The PCAR model seems to show convergence slower than the ICAR model. The blue line is also closer to the red line in the PCAR model and seems more stable than in the ICAR model.

Convergence for the PCAR model is only reached at iterations 1986-3470, which is slower than the ICAR model which reached convergence at iterations 501-1000. From Table 3.27, it may be seen that stability is reached at iterations 5946-11390 and thus the burn-in period is again set to $N_B = 12000$. This is the same burn-in period as the ICAR model in which stability was reached at iterations 5751-11500.

Table 3.28: An analysis of the burn-in period of the PCAR model.

Province	Percentage
Eastern Cape	0.7048
Free State	0.6042
Gauteng	0.7695
KwaZulu-Natal	0.8801
Limpopo	0.5417
Mpumalanga	0.7727
North-West	1.2028
Northern Cape	1.0193
Western Cape	0.7466

As with the ICAR model, and keeping with the rule of thumb proposed by Woodward (2016), the proportion of the MC error to the standard deviation of the relative risk is calculated. All of these proportions are well below 5% and thus, it can be concluded that the burn-in period of $N_B = 12000$ is sufficient, which is the same conclusion as in the ICAR model.

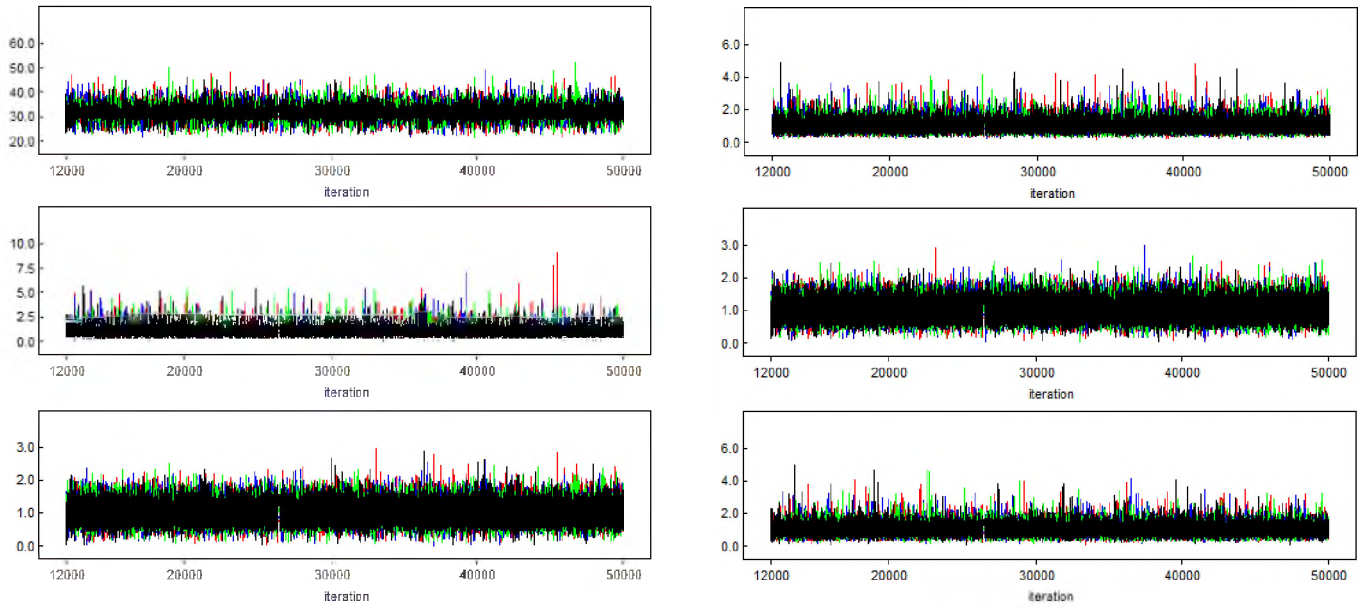


Figure 3.29: The trace plots for multiple chains of the South African data for the PCAR model from top left: Deviance, Eastern Cape, Free State, Gauteng, KwaZulu-Natal and Limpopo.

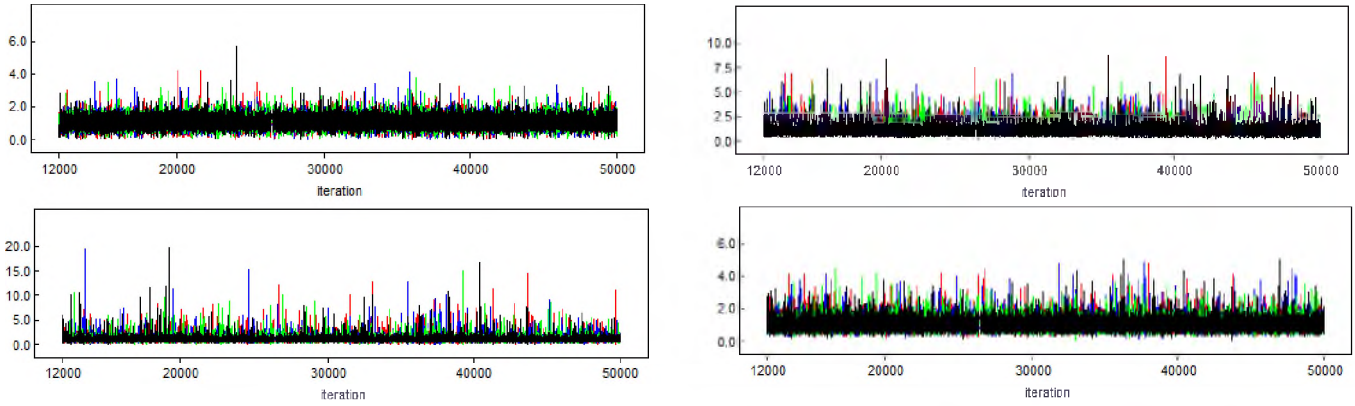


Figure 3.30: The trace plots for multiple chains of the South African data for the PCAR model continued from top left: Mpumalanga, North-West, Northern Cape and Western Cape.

The trace plots in Figures 3.29 and 3.30 show that the relative risk and the deviance show convergence in all provinces. The multiple chains cover each other and start from different initial values thus show convergence.

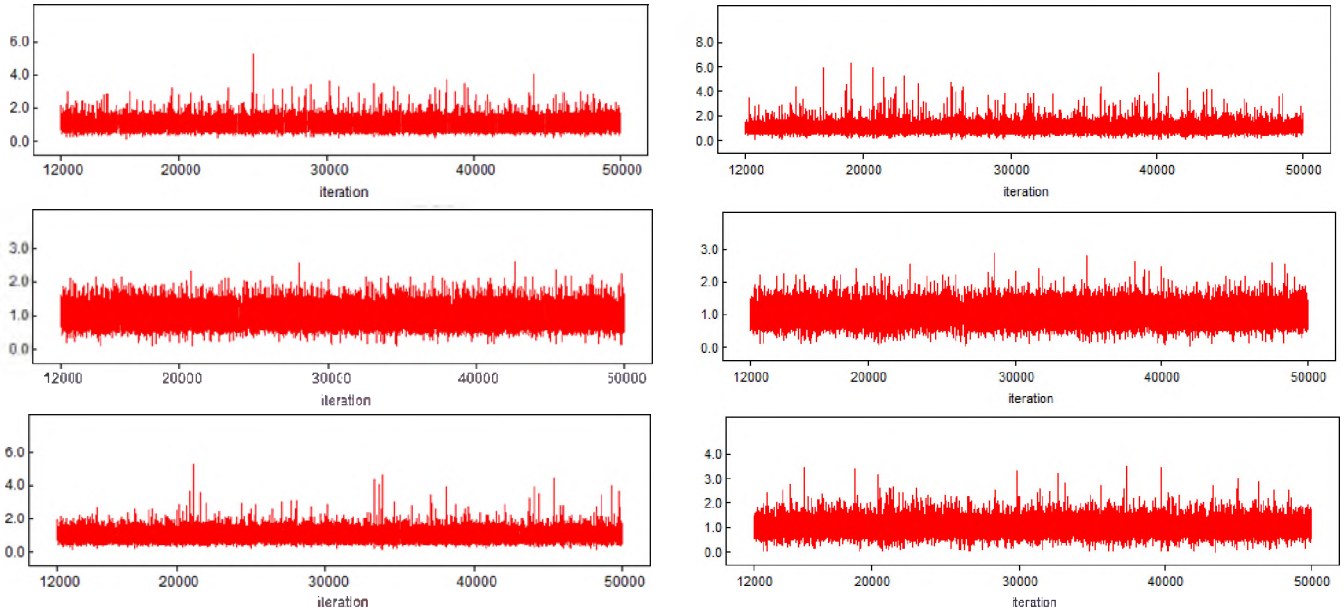


Figure 3.31: The trace plots of the South African data for the PCAR model from top left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo and Mpumalanga.

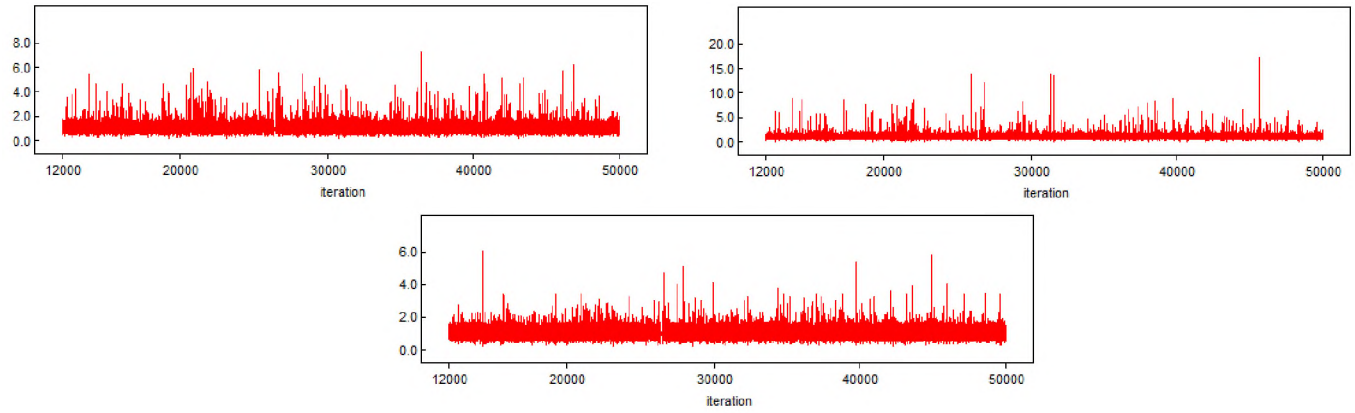


Figure 3.32: The trace plots of the South African data for the PCAR model continued from top left: North-West, Northern Cape and Western Cape.

The PCAR model was run with a single chain of $n_T = 50000$ iterations with a burn-in period of $N_B = 12000$ and produced the trace plots in Figures 3.31 and 3.32. The trace plots show the model converged for the relative risks in all of the provinces.

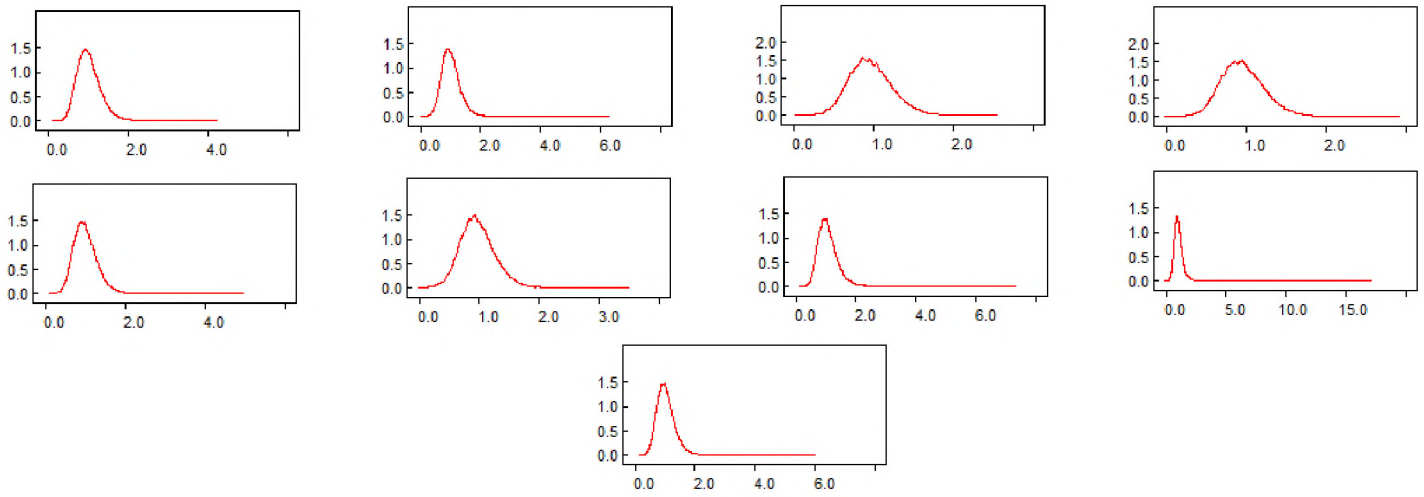


Figure 3.33: The density kernels for the relative risk of the South African data for the PCAR model from top left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape and Western Cape.

The density kernels show a maximum of approximately 1 for the relative risks in all the provinces. All of the provinces show a heavy tail towards the right.

Table 3.29: The model diagnostics of the South African data for the PCAR model.

Statistic	Mean	Standard Deviation	MC Error	95% Confidence Interval	
				2.5%	97.5%
MAPE	1.461	0.399	0.002174	0.7778	2.333
MSPE	3.766	2.345	0.01171	1.111	9.889
pE	0.9449	0.1033	0.002736	0.6222	0.9989
pV	0.02744	0.07497	0.001854	2.37E-4	0.2663
QR60	1.379	0.7671	0.01596	1.039	2.967
pS	0.02768	0.07518	0.001969	2.174E-4	0.25
Saturated Deviance	10.51	1.868	0.01758	6.885	14.99
DIC	32.139				
pD	1.441				
Dhat	29.257				
Dbar	30.698				

The MAPE is the same for both models, while the MSPE is higher in the ICAR model with a difference of 0.008. The quantile ratio, QR60, is 1.379 for the PCAR model and indicates that there is relatively low heterogeneity in the risk across the middle 60% of the provinces. The QR is higher in the PCAR model than in the ICAR model with a difference of 0.058.

The amount of variation attributed to the unobserved spatial factors, pS, is 2.768% of the total variation in the model. This compares to 1.667% in the ICAR model. The variation attributed to the unobserved non-spatial factors, pV, is 2.744% of the total variation in the model which is higher than the 2.647% in the ICAR model. The other 94.49% of the variation in the model is due to the expected values of the provinces.

The DIC is calculated based on an additional $n_A = 20000$ iterations, as in the ICAR model, and resulted in a DIC value of 32.139. This is slightly higher than the ICAR model, which has a DIC value of 32.074. The difference in the saturated deviance between the two models is 0.07 over $n_T = 50000$ iterations, which is not a significant difference.

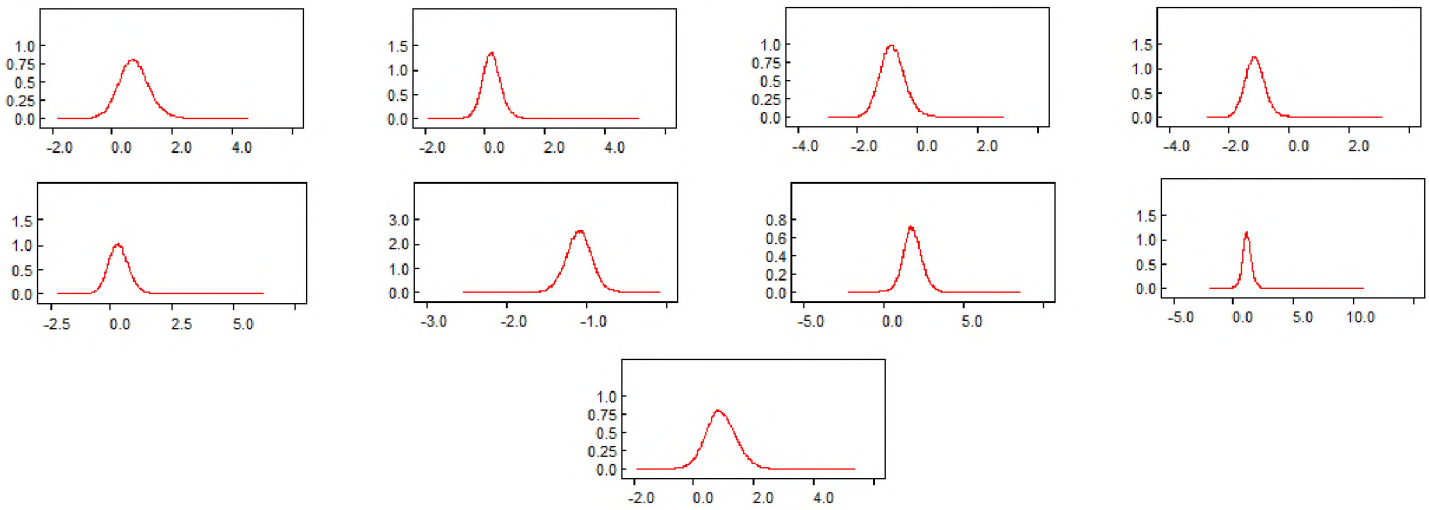


Figure 3.34: The density kernels of the standardised Bayesian residuals of the South African data for the PCAR model from top left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape and Western Cape.

Based on Figure 3.34, Eastern Cape, Free State and Limpopo appear to be approximately centred around 0. KwaZulu-Natal, Mpumalanga, North-West and Western Cape have centres which are further away from 0. An approximate symmetry of the kernel densities is shown in Eastern Cape, Free State, North-West and Western Cape which may indicate a good model fit for these provinces. The symmetry of the other provinces appears to be skewed, with heavy tails occurring at Free State, KwaZulu-Natal, Limpopo and Mpumalanga.

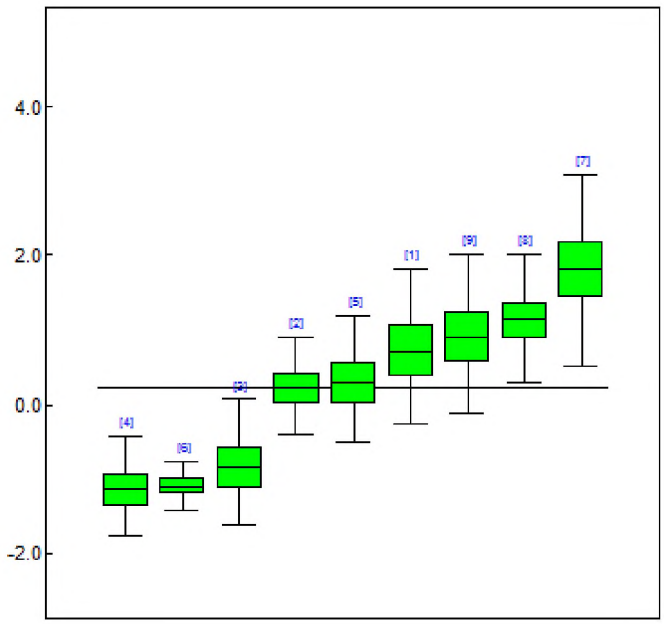


Figure 3.35: The ranked box plots of the standardised Bayesian residuals of the South African data for the PCAR model from left to right: KwaZulu-Natal, Mpumalanga, Gauteng, Free State, Limpopo, Eastern Cape, Western Cape, Northern Cape and North-West.

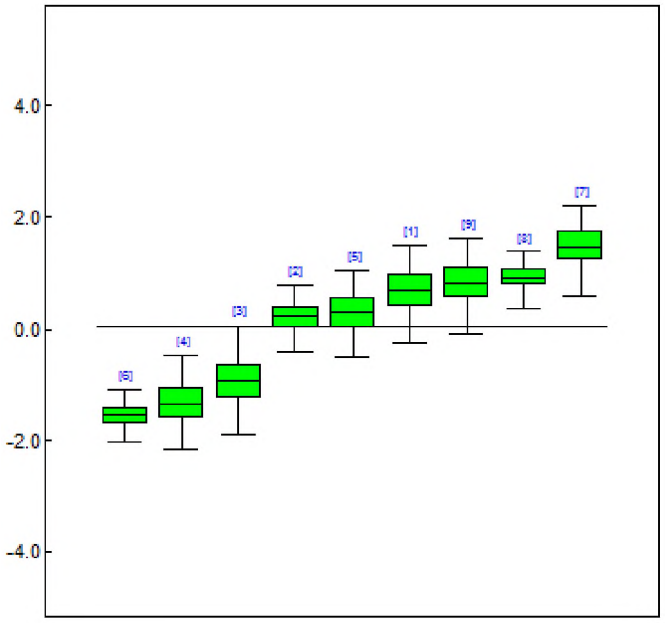


Figure 3.36: The ranked box plots of the deviance residuals of the South African data for the PCAR model from left to right: Mpumalanga, KwaZulu-Natal, Gauteng, Free State, Limpopo, Eastern Cape, Western Cape, Northern Cape and North-West.

The ranked box plot of the standardised Bayesian residuals in Figure 3.35, show that all of the

standardised Bayesian residuals in the PCAR model occur between -2 and 2, thus there are no outliers in the model as with the ICAR model. The North-West is again very close to occurring above 2.

The ranked box plots of the deviance residuals for the PCAR model all occur between -2 and 2 and thus there are no outliers in the PCAR model. The whiskers of the box plot of the deviance residuals of the North-West are shorter than those of the box plot of the standardised Bayesian residuals for the North-West, as was concluded for the ICAR model.

Table 3.30: The statistics from the Bayesian residuals of the South African data for the PCAR model.

Province	Bayesian Residual			95% Confidence Interval	
	Mean	Standard Deviation	MC error	2.5%	97.5%
Eastern Cape	0.9285	0.6277	0.004333	-0.5052	1.911
Free State	0.1575	0.2869	0.002026	-0.4725	0.5889
Gauteng	-1.707	1.032	0.007437	-3.943	0.0962
KwaZulu-Natal	-2.049	0.8697	0.007079	-3.925	-0.5027
Limpopo	0.3108	0.5125	0.003164	-0.8222	1.133
Mpumalanga	-1.224	0.3787	0.002789	-2.05	-0.5658
North-West	1.819	0.4437	0.005242	0.776	2.396
Northern Cape	0.6269	0.1843	0.001834	0.254	0.8335
Western Cape	1.128	0.5761	0.004475	-0.1782	2.011

Table 3.31: The deviance residuals of the PCAR model.

Province	Deviance Residual			95% Confidence Interval	
	Mean	Standard Deviation	MC error	2.5%	97.5%
Eastern Cape	0.654	0.4386	0.002821	-0.2408	1.485
Free State	0.2029	0.2972	0.001971	-0.4014	0.773
Gauteng	-0.9363	0.4918	0.003991	-1.886	0.04419
KwaZulu-Natal	-1.336	0.431	0.003684	-2.158	-0.4744
Limpopo	0.2737	0.3891	0.002281	-0.4986	1.017
Mpumalanga	-1.547	0.2388	0.001986	-2.017	-1.081
North-West	1.463	0.4145	0.004286	0.5604	2.195
Northern Cape	0.8957	0.2719	0.002036	0.3275	1.371
Western Cape	0.8055	0.4293	0.00291	-0.07843	1.616

There are small differences between the Bayesian residuals in the PCAR and ICAR models. The biggest difference between the Bayesian residuals of the two models occurs at the Western Cape with the Bayesian residuals having values of 1.082 and 1.128 for the ICAR and PCAR models, respectively. This is a difference of 0.046, which is not a significant difference. The other provinces all have very small differences between the two models.

As with the Bayesian residuals, the deviance residuals for both models are also very similar.

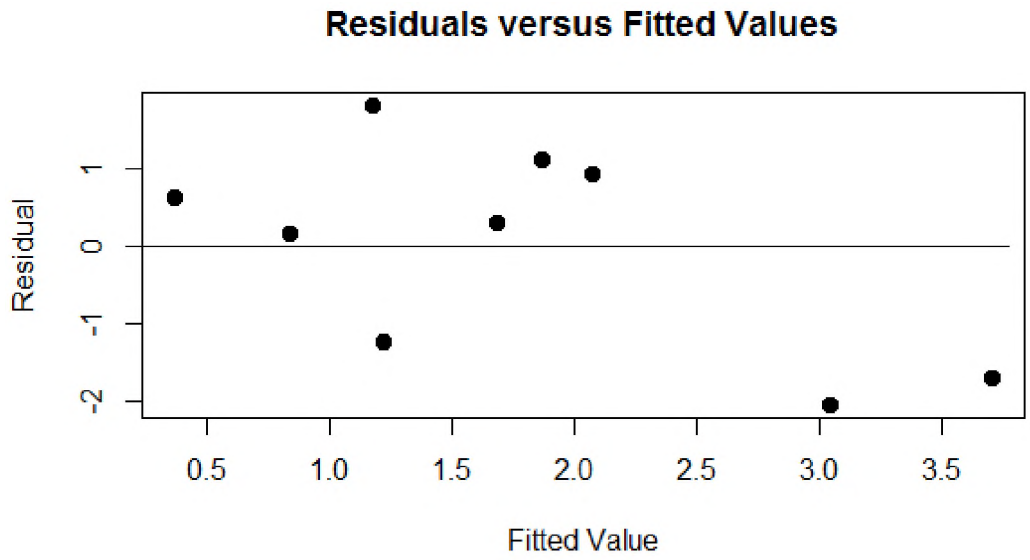


Figure 3.37: The Bayesian residuals versus the fitted values of the PCAR model.

The Bayesian residuals versus fitted values are given in Figure 3.37 and show a fairly random pattern. The pattern seems to change after 2.5 on the fitted values, the figure shows an increase in the spacing after 2.5 than the spacing before 2.5.

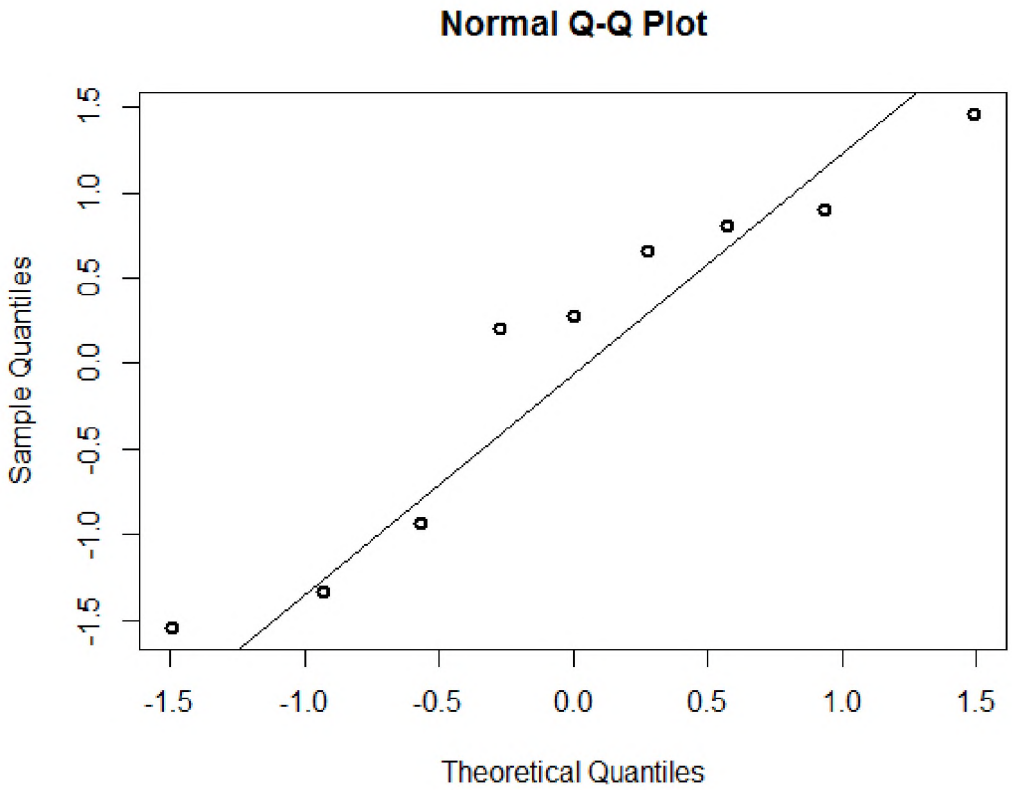


Figure 3.38: Q-Q plot of the deviance residuals of the PCAR model.

The model seems to fit the data adequately as the normal Q-Q plot shows that most of the points are close to the straight line. There is a province at -1.5 which is further away from the straight line than all the other provinces and indicates the model does not fit well to this province. This is different to the Q-Q plot of the ICAR model.

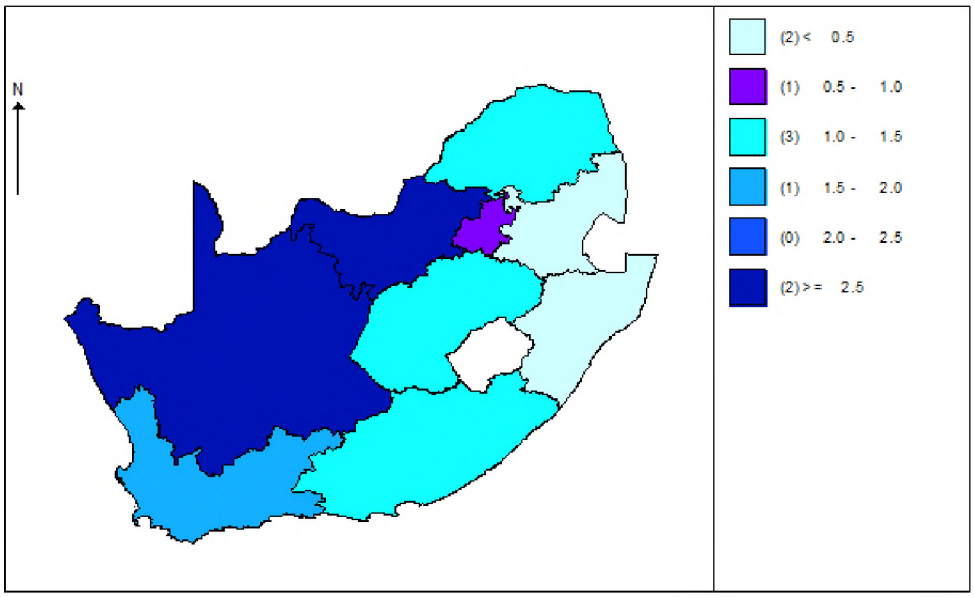


Figure 3.39: The SMR of the South African data for the PCAR model.

The PCAR model produces the same figure as the ICAR model for the SMR. As in the ICAR model, KwaZulu-Natal and Mpumalanga have the lowest SMR values. The highest values of the SMR are again North-West and the Northern Cape. The same conclusions can be drawn from the PCAR model that were drawn from the ICAR model.

Table 3.32: The SMR of the South African data for the PCAR model.

Province	SMR			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Eastern Cape	1.4920	0.007741	0.00004084	1.4770	1.5080
Free State	1.2110	0.008138	0.00004283	1.1970	1.2280
Gauteng	0.5226	0.003944	0.00002082	0.5148	0.5302
KwaZulu-Natal	0.3155	0.006830	0.00003606	0.3019	0.3286
Limpopo	1.1990	0.003775	0.00001991	1.1920	1.2070
Mpumalanga	0.0007	0.025300	0.00013340	-0.0522	0.0467
North-West	2.7560	0.051110	0.00026930	2.6610	2.8600
Northern Cape	2.9100	0.179800	0.00094090	2.6070	3.3080
Western Cape	1.6560	0.011450	0.00006040	1.6340	1.6790

The SMR values in Table 3.32 give almost the same means as the ICAR, with the only difference in the SMR value of Mpumalanga. In the ICAR model, the mean SMR value at Mpumalanga is 0.0007581

whereas in the PCAR model the value is 0.0007134. The mean SMR value at Mpumalanga is slightly higher than the PCAR model with a difference of only 0.0000447. Other slight differences do occur in the standard deviations and MC errors between the two models. The standard deviations are slightly different for the two models, with the PCAR model having a slightly lower standard deviation for all the provinces. The MC error, as with the standard deviation, is slightly lower for the PCAR model than for the ICAR model.

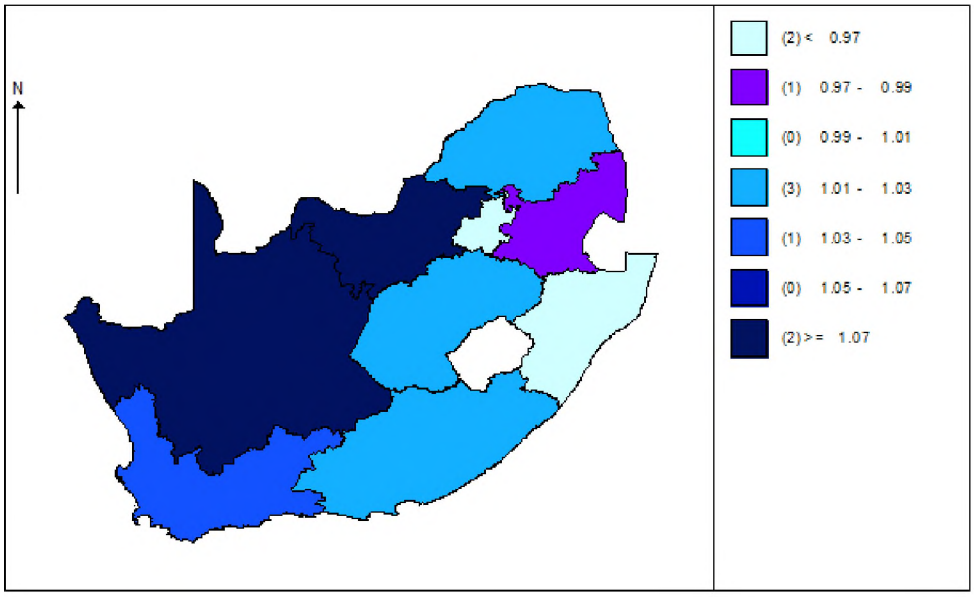


Figure 3.40: The relative risk of the South African data for the PCAR model.

Table 3.33: The relative risk of the South African data for the PCAR model.

Province	Relative Risk			95% Confidence Interval	
	Mean	Standard Deviation	MC Error	2.5%	97.5%
Eastern Cape	1.0300	0.3130	0.002206	0.5434	1.712
Free State	1.0240	0.3522	0.002128	0.4984	1.805
Gauteng	0.9696	0.2720	0.002093	0.4994	1.560
KwaZulu-Natal	0.9632	0.2786	0.002452	0.4759	1.547
Limpopo	1.0130	0.3085	0.001671	0.5267	1.692
Mpumalanga	0.9757	0.3032	0.002342	0.4595	1.628
North-West	1.0830	0.4018	0.004833	0.5536	2.017
Northern Cape	1.0750	0.5235	0.005336	0.4881	2.136
Western Cape	1.0360	0.3204	0.002392	0.5457	1.760

The relative risk is slightly higher for the ICAR model than the PCAR model in provinces Gauteng and the Western Cape. All the other provinces have a slightly lower relative risk in the ICAR model. The highest difference is 0.023. The same can be concluded for the PCAR model as was concluded

for the ICAR model, in that the relative risk of contracting and dying from acute pericarditis in South Africa in 2014 has a relatively low risk.

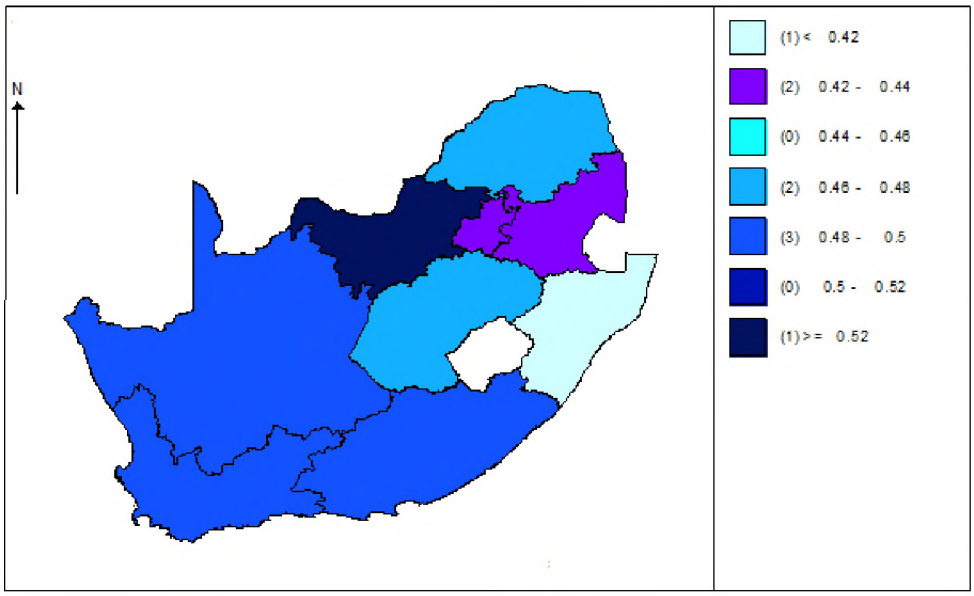


Figure 3.41: The posterior exceedance probability of the South African data for the PCAR model.

Table 3.34: The posterior exceedance probability of the South African data for the PCAR model.

Posterior Expected Exceedance Probability			
Province	**Mean**	**Standard Deviation**	**MC Error**
Eastern Cape	0.4887	0.4999	0.002814
Free State	0.4750	0.4994	0.002656
Gauteng	0.4257	0.4944	0.003089
KwaZulu-Natal	0.4146	0.4927	0.003332
Limpopo	0.4693	0.4991	0.002616
Mpumalanga	0.4332	0.4955	0.003117
North-West	0.5265	0.4993	0.003634
Northern Cape	0.4914	0.4999	0.002407
Western Cape	0.4943	0.5000	0.002807

The PP for the PCAR model shows that there is a slight difference in the PCAR model compared to the ICAR model. The biggest difference occurs at the Western Cape where the ICAR model has a slightly higher value than the PCAR model with the difference being 0.0141. There are three provinces where the PP for the ICAR model is lower than the PP for the PCAR model, these occur at Gauteng, KwaZulu-Natal and Mpumalanga. The standard deviation for the PP in the ICAR model is lower or equal to the standard deviation for the PP in the PCAR model in 7 provinces. There are slight differences in the standard deviations and MC errors in the two models.

The PP shows that there is no usual or unexpected clustering of the disease in the PCAR model when the exceedance probability, $\hat{P}$, of interest is set at 0.9.

The confidence intervals for the PP are not given in Table 3.34 as all of the confidence intervals have a lower bound of 0 and an upper bound of 1.0.

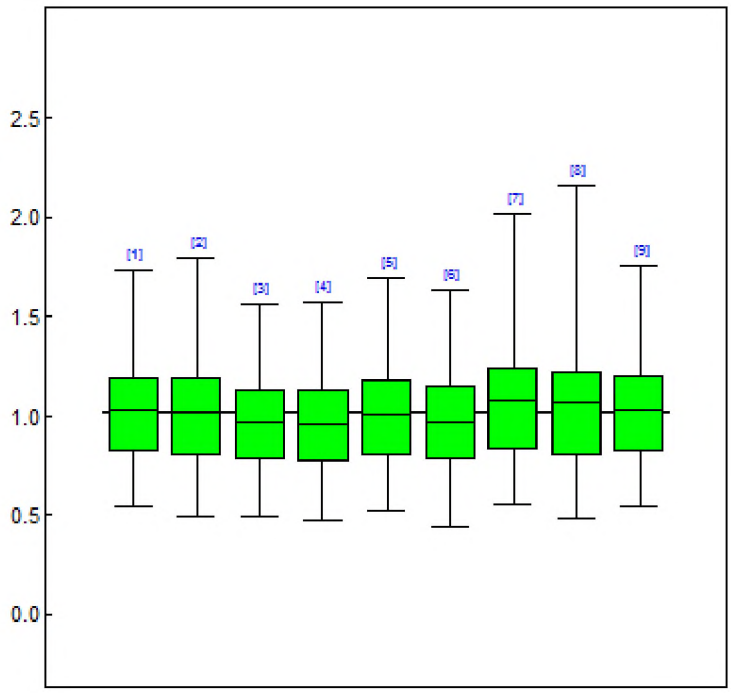


Figure 3.42: The box plot of the relative risk of the South African data for the PCAR model from the left: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape and Western Cape.

The box plot for the relative risk of the PCAR model is centred around 1 with North-West, Western Cape and Northern Cape slightly higher than all the other provinces. The Northern Cape shows a heavier tail than all the other provinces which is expected because of the high standard deviation of the relative risk in the province.

Chapter 4

Conclusion

4.1 Concluding Remarks

Oral cancer in Georgia in 2004 was modelled for 3 models. The best model, based on the DIC, MAPE and MSPE, was model 1 which had a conjugate hyperprior distribution for the precision of the UH component. The parameters of the conjugate hyperprior distribution were calculated based upon approximately 30% of the data in the model. The difference between models 1 and 2 was not significant as the difference in the DIC values of the two models was only 2.664. We considered a difference of more than 10 in the DIC to be a significant difference. There was, however, differences between the results of the two models even though the difference in the DIC was small. The difference in DIC was significant between models 1, 2 from model 3. This difference in DIC from model 3 was 29.011 and 31.675 for models 2 and 1, respectively. The results produced for models 1 and 2 were comparable, and both models fitted the data adequately whereas model 3 had a poor fit to the data and produced significantly different results to those of models 1 and 2. The results of model 3 were only used for illustrative purposes as the model fit was poor.

The sensitivity of the hyperprior for the precision of the UH component was illustrated by changing from a hyperprior which allows for some flexibility and estimation in the model to a fixed parameter for the precision which does not allow flexibility or estimation in the model. It was therefore important to select the hyperprior distributions for the precision of the UH component carefully. The same can be concluded with the precision of the ICAR prior. This hyperprior for the precision of the ICAR prior was also sensitive to changes and should be selected with care.

The variation in the model due to the unobserved non-spatial factors and unobserved spatial factors may be changed by manipulating the hyperpriors for the precision, as was illustrated in models 1 and 2. It was often the case that either the UH component was dominant over the CH component or vice versa, although the dominance was not known before the model was run and could only be seen after the model was run. The CH component was more dominant over the UH component in model 2 whereas the UH component was more dominant over the CH component in model 1.

The fixed parameter in model 3 resulted in a very high proportion of the total variation in the model due to the unobserved non-spatial factor. This was higher than the variation due to the expected values and is not desirable in any model. The variation due to the unobserved non-spatial factor may be reduced by changing the value of the fixed parameter.

The ICAR and PCAR models for the South African data show that there was no significant difference between the ICAR and PCAR models based on the model diagnostics. The DIC for each model was 32.074 and 32.139 for the ICAR and PCAR models, respectively. The difference in DIC was only 0.065 which was not large enough to consider a significant difference between the two models.

The model diagnostics also show that there was a slightly higher proportion of the total variation due to the expected values in the ICAR model, a difference of 1.20% from the PCAR model. The proportion of the total variation due to the unobserved spatial factors was higher in the PCAR model with 2.77% compared with the 1.67% in the ICAR model, a difference of 1.10%. The remaining proportion of the total variation was due to the unobserved non-spatial factors which was higher in the PCAR model than in the ICAR model. Although these proportions may be changed by changing the parameters for the hyperprior distributions for both models.

The other results from both models were very similar apart from very slight but insignificant differences. Other differences occurred in the standard deviation of statistics that were calculated for both models, however, these differences change from variable to variable and it was difficult to conclude which model results had a lower or higher standard deviation than the other model.

The results from the ICAR and PCAR models showed that it was possible to use either model and still obtain very similar results for the PCAR and ICAR models for acute pericarditis in South Africa in 2014. Both the ICAR and PCAR models also showed that there was a low risk of contracting and dying from acute pericarditis in all provinces in 2014. The ICAR model seems to be the more popular model among researchers. We found the disadvantages of the PCAR as:

- The implementation of the PCAR model was more complex than the ICAR model.
- The WinBUGS® code for the PCAR model was clumsy.
- The PCAR model had a longer computation time than the ICAR model. This became evident when the number of parameters and number of regions in the model increased. Some of the computations may be done in RStudio® and the result of these computations may be inserted into WinBUGS® as data as a temporary fix for the longer computation time.
- The PCAR model only worked for models with a small number of regions and a small number of parameters. The PCAR model could not be run for the Georgia data as a result.
- The choice of initial values in the PCAR model may lead to a trap error in WinBUGS®, thus the specification of initial values in the PCAR model was more difficult than in the ICAR model.

- An incorrect specification of the bounds of the parameter ρ, the amount of spatial dependence, in the PCAR model resulted in incorrect results.

4.2 Future Research

The sample size of each region should be increased and previous year's data for each region should be included in the model. This would enable an investigation into the trend of the relative risk in each region over time. The increase in sample size would also help the model produce more reliable results. The coordinates of each observation would also be helpful in analysing spatial correlation as well as the identification of clusters in the data. Cluster analysis should be performed on the data to check that there are no regions with unusual risk.

The complexity of diseases made it difficult to model a contagious disease as the independence of the data would be broken. The likelihood is based on the foundation that the data is independent thus an investigation into the methods and techniques required to perform such an analysis would be very interesting. Another issue, requiring further investigation, would be the analysis of a disease which results in a higher number of deaths. The Poisson distribution is used when there are a small number of deaths as a large number of deaths create a lot of overdispersion. It would be interesting to investigate which distribution would fit the data well in that case.

Defining a model which can handle multiple diseases at once and a mapping technique to map the output of such a model.

Investigating other applications of disease mapping in financial markets and econometric applications may also be a possibility.

References

Ahmad, O. B., Boschi-Pinto, C., Lopez, A. D., Murray, C. J., Lozano, R., & Inoue, M. (2001). Age standardization of rates: a new WHO standard. *GPE Discussion Paper Series: No. 31*.

Berk, R. & MacDonald, J. M. (2008). Overdispersion and Poisson regression. *Journal of Quantitative Criminology*, 24(3), 269–284.

Bernardinelli, L., Clayton, D. G., & Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, 14, 2411–2431.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2), 113–147.

Besag, J. & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4), 733–746.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.

Bivand, R. & Lewin-Koh, N. (2013). Maptools: Tools for reading and handling spatial objects. *R package version 0.8*, 27.

Bivand, R. & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, 63(18), 1–36.

Bolstad, W. (2004). *Introduction to Bayesian Statistics*. John Wiley and Sons, Inc.

Box, G. E. P. & Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley and Sons, Inc., Wiley Classics Library edition.

Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51(3/4), 481–483.

Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 69–100.

Brooks, S. P., Giudici, P., & Philippe, A. (2003). Nonparametric convergence assessment for MCMC model selection. *Journal of Computational and Graphical Statistics*, 12, 1–22.

Carneiro, I., Howard, N., & Bailey, L. (2011). *Introduction to Epidemiology*. Understanding Public Health. Open University Press.

Casella, G. & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), 167–174.

Centers for Disease Control and Prevention (CDC) (2012). *Principles of Epidemiology in Public Health Practice*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), Office of Workforce and Career Development, Atlanta, Georgia 30333, Third edition.

Coggon, D., Rose, G., & Barker, D. (2009). *Epidemiology for the Uninitiated*. Wiley, Fifth edition.

Cowles, M. K. & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434), 883–904.

Cressie, N. & Chan, N. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, 84(406), 393–401.

Curtin, L. R. & Klein, R. J. (1995). *Direct Standardization (Age-Adjusted Death Rates)*. Healthy People 2000 Statistical Notes, no. 6-revised. Hyattsville, Maryland: National Center for Health Statistics.

Datta, G. S. & Sweeting, T. J. (2005). Probability matching priors. In *Handbook of Statistics*, Dey, D. K. & Rao, C. R. (eds), volume 25: Bayesian Thinking: Modeling and Computation (pp. 91–114). Elsevier.

Davies, J., Wilson, B., & Knowelden, J. (1962). Cancer incidence of the African population of Kyadondo (Uganda). *Lancet*, ii, 328–330.

Department of Health and Human Services (2002). Standards for privacy of individually identifiable health information. *Rules and Regulations: Federal Register*, 67(157), 53182–53273.

Diaconis, P. & Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1), 36–61.

DiMaggio, C. (2012). *Bayesian Hierarchical Approaches to Spatial Analysis of Injury and Disaster Data*. PhD thesis, Columbia University, Departments of Anesthesiology and Epidemiology, 116th and Broadway, New York, NY 10027.

DLA Cliffe Dekker Hofmeyr (2012). *Data protection and privacy alert*. DLA Cliffe Dekker Hofmeyr, 1 Protea Place, Sandton, Johannesburg, 2196, Private Bag X40, Bermore, 2010, South Africa, August edition. Retrieved from https://www.cliffedekkerhofmeyr.com/export/sites/cdh/en/news/publications/2012/technology/downloads/Data-Protection-and-Privacy-Alert—22-August-2012.pdf.

Doll, R. & Cook, P. (1967). Summarizing indices for comparison of cancer incidence data. *International Journal of Cancer*, 2(3), 269–279.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30, 609–628.

Douglas, E., Vogel, R., & Kroll, C. (2000). Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology*, 240, 90–105.

Edwards, R. G. & Sokal, A. D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review D*, 38, 2009–2012.

Efron, B. & Morris, C. (1973). Stein's estimation rule and its competitors–an empirical Bayes approach. *Journal of the American Statistical Association*, 68(341), 117–130.

Elliott, P. & Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives*, 112(9), 998–1006.

Farr, W. (1859). *Letter to the Registrar General*. General Registrar Office, London.

Fleiss, J. L., Levin, B., & Paik, M. C. (1981). *Statistical Methods for Rates and Proportions*. John Wiley & Sons, Inc., Third edition.

Gelfand, A. E. & Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85(1), 1–11.

Gelfand, A. E., Sahu, S. K., & Carlin, B. P. (1995). Efficient parameterizations for normal linear mixed models. *Biometrika*, 82, 479–488.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press: Taylor & Francis Group, Third edition.

Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.

Ghosh, M., Mergel, V., & Liu, R. (2011). A general divergence criterion for prior selection. *Annals of the Institute of Statistical Mathematics*, 63(1), 43–58.

Giuffré, O., Graná, A., Roberta, M., & Corriere, F. (2011). Handling underdispersion in calibrating safety performance function at urban, four-leg, signalized intersections. *Journal of Transportation Safety & Security*, 3(3), 174–188.

Greenland, S., Schwartzbaum, J. A., & Finkle, W. D. (2000). Problems due to small samples and sparse data in conditional logistic regression analysis. *American Journal of Epidemiology*, 151(5), 531–539.

Health Insurance Portability and Accountability Act (1996). *Public Law 104-191*. United States Congress, Administrative Counsel, Ford House Office Building, Room 217, Washington, DC 20515.

Hellinger, E. (1909). Neue begründung der theorie quadratischen formen von unendlichen vielen veränderlichen. *Journal Für Reine und Angewandte Mathematik*, 136, 210–271.

Hijmans, R. (2016). *Spatial autocorrelation*. Retrieved from http://rspatial.org/analysis/rst/3-spauto.html.

Imazio, M., Cecchi, E., Demichelis, B., Ierna, S., Demarie, D., Ghisio, A., Pomari, F., Coda, L., Belli, R., & Trinchero, R. (2007). Indicators of poor prognosis of acute pericarditis. *Circulation*, 115(21), 2739–2744.

Jones, M. & Swerdlow, A. (1998). Bias in the standardized mortality ratio when using general population rates to estimate expected number of deaths. *American Journal of Epidemiology*, 148(10), 1012–1017.

Jordan, M. I. (2010). *Jeffreys Priors and Reference Priors*. University of California, Statistics Department, 427 Evans Hall, Berkeley, California, USA 94720-3860.

Julious, S. A., Nicholl, J., & George, S. (2001). Why do we continue to use standardized mortality ratios for small area comparisons? *Journal of Public Health Medicine*, 23(1), 40–46.

Kass, R. E. & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435), 1343–1370.

Keiding, N. (1987). The method of expected number of deaths 1786-1886-1986, correspondent paper. *International Statistical Review / Revue Internationale de Statistique*, 55(1), 1–20.

Keiding, N. & Clayton, D. (2014). Standardization and control for confounding in observational studies: a historical perspective. *Statistical Science*, 29(4), 529–558.

Kelsall, J. & Wakefield, J. (1999). Discussion of "Bayesian models for spatially correlated disease and exposure data". In *Bayesian Statistics 6*, Bernardo, J. M., Berger, J. O., Dawid, A. P. & Smith, A. F. M. (eds). Oxford University Press, Oxford.

King, G. (1998). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Political Analysis. University of Michigan Press.

Kirita, T. & Omura, K. (2015). *Oral Cancer: Diagnosis and Therapy*. Springer Japan.

Kokonendji, C. C., Mizére, D., & Balakrishnan, N. (2008). Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference*, 138, 1287–1296.

Künsch, H. R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika*, 74(3), 517–524.

Lange, R. A. & Hillis, L. D. (2004). Acute pericarditis. *New England Journal of Medicine*, 351(21), 2195–2202.

Lawson, A. B. (2003). Introduction to Bayesian mapping methods. Conference Slides, Medical University of South Carolina, 135 Cannon St Suite 303 MSC835, Charleston South Carolina, 29401-8350, USA.

Lawson, A. B. (2009). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. CRC Press.

Lawson, A. B. (2014). *Disease Clustering*. Medical University of South Carolina, Department of Medicine, Medical University of South Carolina, 135 Cannon St Suite 303 MSC835, Charleston South Carolina, 29401-8350, USA.

Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. (2003). *Disease Mapping with WinBUGS and MLwiN*. John Wiley & Sons, Inc.

Lee, P. M. (2004). *Bayesian Statistics: An Introduction*. Arnold Publishers, Third edition.

Lesaffre, E. & Lawson, A. B. (2012). *Bayesian Biostatistics*. John Wiley & Sons, Ltd.

Liang, K. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22.

Lunn, D. J., Jackson, C., Best, N. G., Thomas, A., & Spiegelhalter, D. J. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press: Taylor & Francis Group.

Lunn, D. J., Thomas, A., Best, N. G., & Spiegelhalter, D. J. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, (10), 325–337.

MacMahon, B. & Pugh, T. F. (1970). *Epidemiology: principles and methods*. Boston: Little Brown & Co. Published in Great Britain by J. & A. Churchill, London.

Mark, B. & Turner, T. R. (1992). Approximating point process likelihoods with GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1), 31–38.

Marshall, R. J. (1991). Mapping disease and mortality rates using empirical Bayes estimators. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 40(2), 283–294.

Montgomery, D., Peck, E., & Vining, G. (2015). *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), 17–23.

Naing, N. N. (2000). Easy way to learn standardization: Direct and indirect methods. *Malaysian Journal of Medical Sciences*, 7(1), 10–15.

Neison, F. G. P. (1844). On a method recently proposed for conducting inquiries into the comparative sanatory condition of various districts, with illustrations, derived from numerous places in Great Britain at the period of the last census. *Journal of the Royal Statistical Society, London*, 7, 40–68.

Ogle, W. M. (1892). Proposal for the establishment and international use of a standard population. *Bulletin de l'Institut International de Statistique, tome VI*, livre 1, 83–85.

Osei, P. P. (2010). *Statistical Methods for Disease Mapping*. PhD thesis, African Institute for Mathematical Sciences (AIMS), 6 Melrose Road, Muizenberg, Cape Town, South Africa 7950.

Paradis, E. (2017). Moran's autocorrelation coefficient in comparative methods.

Pickle, L. W. (2002). Spatial analysis of disease. *National Cancer Institute*, 113, 113–150.

Pickle, L. W. & White, A. A. (1995). Effect of the choice of age-adjustment on maps of death rates. *Statistics in Medicine*, 14, 615–627.

Press, S. (2009). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

Quantum GIS Development Team (2016). *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project., 2.14 LTR Essen edition.

Raubenheimer, L. & Van der Merwe, A. (2014). Bayesian estimation for linear functions of Poisson rates. *Communications in Statistics - Theory and Methods*, 43(19), 4025–4045.

Reese, T. (2010). *Colonial Georgia: A Study in British Imperial Policy in the Eighteenth Century*. University of Georgia Press.

Robert, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. Springer New York.

Robert, C. P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag New York, Inc., Second edition.

Rosenberg, H. M. & Anderson, R. N. (1998). Report of the second workshop on age adjustment. *Vital and Health Statistics. Series 4: Documents and Committee Reports*, (30), 1–37.

Ross, S. M. (2010). *Introduction to Probability Models*. Academic Press, Tenth edition.

Rothman, K. J. (2012). *Epidemiology: An Introduction*. Oxford University Press, Second edition.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern Epidemiology*. Wolters Kluwer Health/-Lippincott Williams & Wilkins.

RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, Massachusetts, USA.

Sankoh, O., Sharrow, D., Herbst, K., Kabudula, C. W., Alam, N., Kant, S., Ravn, H., Bhuiya, A., Vui, L. T., Darikwa, T., Gyapong, M., Jasseh, M., Kim, N. C. T., Abdullah, S., Crampin, A., Ojal, J., Owusu-Agyei, S., Odhiambo, F., Urassa, M., Streatfield, K., Shimada, M., Sacoor, C., Beguy, D., Derra, K., Wak, G., Delaunay, V., Sie, A., Soura, A., Diallo, D., Wilopo, S., Masanja, H., Bonfoh, B., Phuanukoonnon, S., & Clark, S. J. (2014). The INDEPTH standard population for low- and middle-income countries, 2013. *Global Health Action*, 7(23286).

Segi, M. (1960). *Cancer Mortality for Selected Sites in 24 Countries (1950-57), Issue 5*. Department of Public Health, Tohoku University School of Medicine, Sendi, Japan.

Selvi, V. S. & Nishanthi, M. (2012). Mathematical applications of queueing theory in call centers. *International Journal of Science and Engineering Research*, 3.

South African Government (1992). *Government Gazette: Births and Deaths Registration Act, 1992 (Act No. 51 of 1992).* Republic of South Africa, Cape Town, South Africa. Volume 323.

South African Government (2013). *Government Gazette: Protection of Personal Information Act, 2013 (Act No. 4 of 2013).* Republic of South Africa, Cape Town, South Africa.

South African Law Reform Commission (2005). *Privacy and Data Protection.* South African Law Reform Commission, Private Bag X668, Pretoria, South Africa 0001.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64, 583–639.

Spiegelhalter, D. J., Thomas, A., & Best, N. G. (2003). *WinBUGS Version 1.4 User Manual.* Imperial College and Medical Research Council, United Kingdom.

Stampfer, M. J. & Colditz, G. A. (1991). Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Preventive medicine*, 20(1), 47–63.

Statistics South Africa (2014). *Mid-year Population estimates.* Statistics South Africa, Private Bag X44, Pretoria, South Africa 0001.

Statistics South Africa (2015). *Mortality and causes of death in South Africa, 2014: Findings from death notification.* Statistics South Africa, Private Bag X44, Pretoria, South Africa 0001.

Stern, H. S. & Cressie, N. A. C. (1999). Inference for extremes in disease mapping. In *Disease Mapping and Risk Assessment for Public Health, Chapter 5*, Lawson, A. B., Biggeri, A., Boehning, D., Lesaffre, E., Viel, J. F. & Bertollini, R. (eds). New York: Wiley.

Sudhakar, A. (2009). History of cancer, ancient and modern treatment methods. *Journal of cancer science & therapy*, 1(2), 1–4.

Sun, D. (1997). A note on informative priors for Weibull distributions. *Journal of Statistical Planning and Inference*, 61, 319–338.

Tango, T. (2010). *Statistical Methods for Disease Clustering.* Statistics for Biology and Health. Springer New York.

Waller, L. A. & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

West, B. (2008). *Poisson Distribution.* McGill University, McConnell Engineering Bldg, Room 318, Montreal, Quebec, Canada, H3A 0E9.

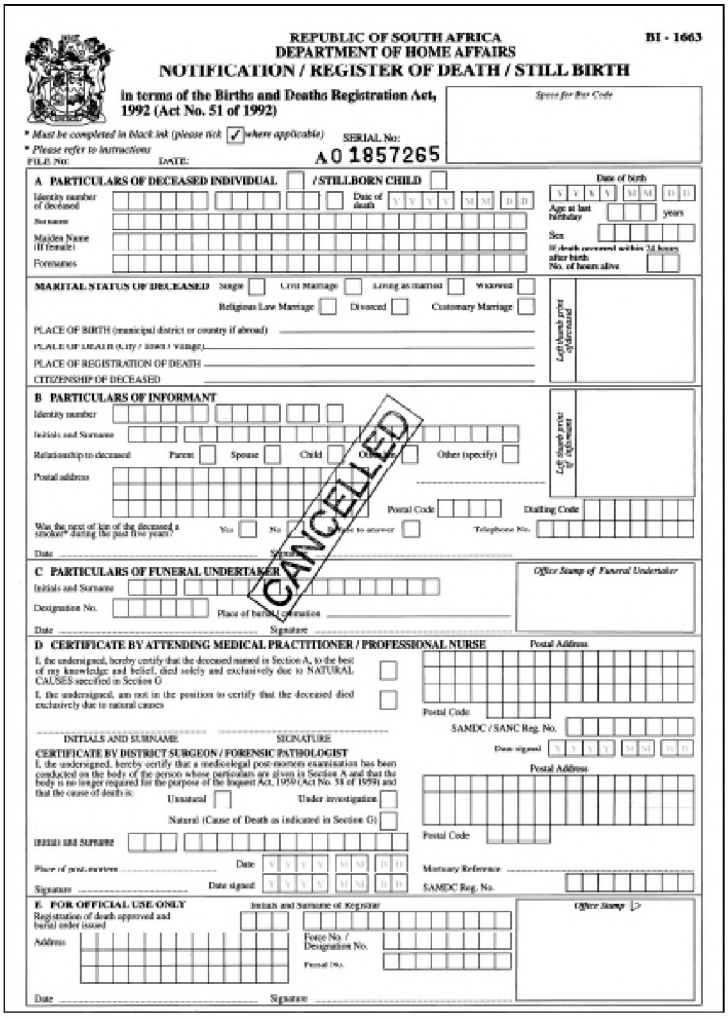
Woodward, P. (2016). *Bayesian Analysis Made Simple: AN Excel GUI for WinBUGS*. Chapman & Hall/CRC Biostatistics Series. CRC Press.

World Health Organization (2017). *Epidemiology*. Retrieved from http://www.who.int/topics/epidemiology/en/.

Writing Group for the Women's Health Initiative Investigators (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomized controlled trial. *JAMA*, 288(3), 321–333.

Yang, Z., Hardin, J. W., Addy, C. L., & Vuong, Q. H. (2007). Testing approaches for overdispersion in Poisson regression versus the generalized Poisson model. *Biometrical Journal*, 49(4), 565–584.

Appendix A: Certificates



REPUBLIC OF SOUTH AFRICA
DEPARTMENT OF HOME AFFAIRS
NOTIFICATION / REGISTER OF DEATH / STILL BIRTH

BI - 1663

in terms of the Births and Deaths Registration Act, 1992 (Act No. 51 of 1992)

SERIAL No: A 0 1857265

A PARTICULARS OF DECEASED INDIVIDUAL / STILLBORN CHILD

B PARTICULARS OF INFORMANT

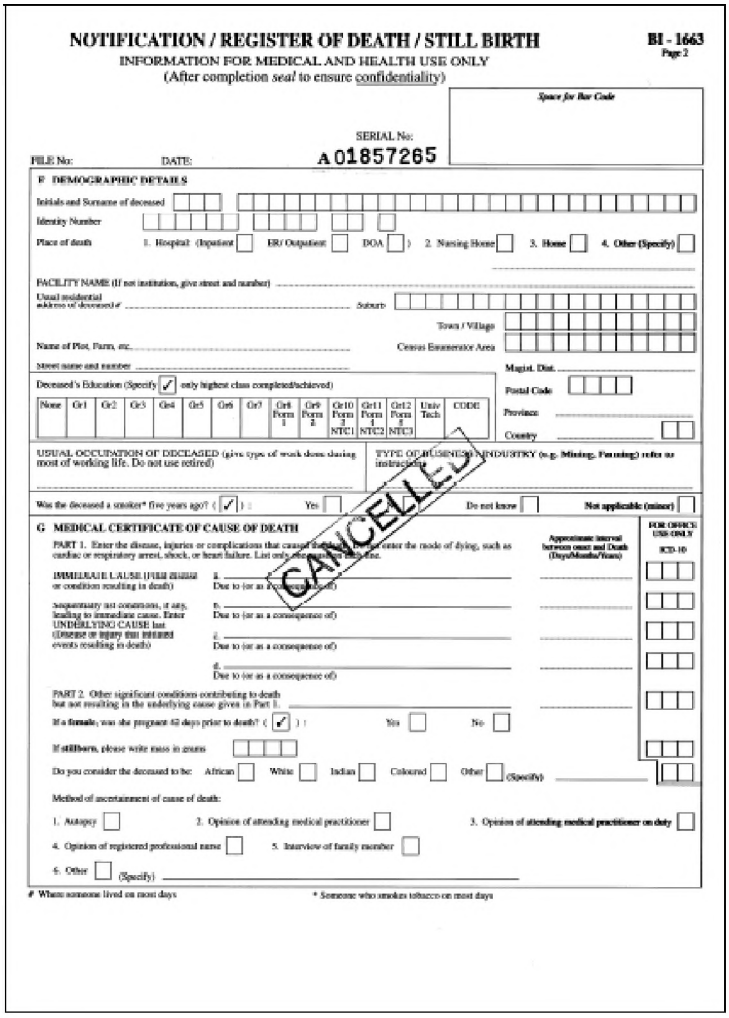
C PARTICULARS OF FUNERAL UNDERTAKER

D CERTIFICATE BY ATTENDING MEDICAL PRACTITIONER / PROFESSIONAL NURSE

CERTIFICATE BY DISTRICT SURGEON / FORENSIC PATHOLOGIST

E FOR OFFICIAL USE ONLY

CANCELLED



NOTIFICATION / REGISTER OF DEATH / STILL BIRTH

BI - 1663
Page 2

INFORMATION FOR MEDICAL AND HEALTH USE ONLY
(After completion *seal* to ensure confidentiality)

SERIAL No: A 01857265

F DEMOGRAPHIC DETAILS

G MEDICAL CERTIFICATE OF CAUSE OF DEATH

CANCELLED

Figure A.1: Death notification for the capturing of data.

Appendix B: Data

B.1 Georgia Data

Table B.1: The Georgia population by county in 2004.

County	Population	County	Population	County	Population
Appling	17740	Clinch	6893	Glynn	71475
Atkinson	7985	Cobb	640346	Gordon	49426
Bacon	10283	Coffee	39383	Grady	23840
Baker	4024	Colquitt	42987	Greene	15330
Baldwin	45967	Columbia	102934	Gwinnett	687468
Banks	16250	Cook	16268	Habersham	38491
Barrow	55032	Coweta	104089	Hall	156385
Bartow	87834	Crawford	12784	Hancock	9673
Ben Hill	17106	Crisp	22320	Haralson	27621
Berrien	17399	Dade	16028	Harris	27180
Bibb	154159	Dawson	18831	Hart	23966
Bleckley	12018	Decatur	27763	Heard	11273
Brantley	16222	De Kalb	666204	Henry	159971
Brooks	16127	Dodge	20319	Houston	123723
Bryan	26061	Dooly	12848	Irwin	9402
Bulloch	60832	Dougherty	94596	Jackson	48288
Burke	22867	Douglas	107377	Jasper	12588
Butts	21706	Early	11800	Jeff Davis	13433
Calhoun	6306	Echols	3926	Jefferson	16988
Camden	46096	Effingham	43674	Jenkins	8490
Candler	10039	Elbert	20693	Johnson	9613
Carroll	99774	Emanuel	21579	Jones	26537
Catoosa	58866	Evans	10889	Lamar	16567
Charlton	11378	Fannin	21873	Lanier	7824
Chatham	240818	Fayette	99443	Laurens	46329
Chattahoochee	15515	Floyd	94014	Lee	25965
Chattooga	25537	Forsyth	129639	Liberty	63254
Cherokee	173105	Franklin	21430	Lincoln	8284
Clarke	109752	Fulton	809481	Long	11355
Clay	3233	Gilmer	26175	Lowndes	96510
Clayton	255322	Glascock	2720	Lumpkin	24916

Table B.2: The Georgia population by county in 2004 continued.

County	Population	County	Population	County	Population
McDuffie	21256	Pulaski	10646	Toombs	26166
McIntosh	12182	Putnam	19810	Towns	9872
Macon	14304	Quitman	2457	Treutlen	6949
Madison	26859	Rabun	15757	Troup	62001
Marion	7637	Randolph	7726	Turner	9310
Meriwether	22563	Richmond	196883	Twiggs	10266
Miller	6135	Rockdale	76577	Union	19378
Mitchell	23520	Schley	4070	Upson	27514
Monroe	23648	Screven	15047	Walker	64228
Montgomery	9040	Seminole	9065	Walton	70334
Morgan	16604	Spalding	60588	Ware	34991
Murray	39465	Stephens	25314	Warren	6195
Muscogee	185057	Stewart	5534	Washington	21088
Newton	80603	Sumter	32902	Wayne	28227
Oconee	27904	Talbot	6775	Webster	2549
Oglethorpe	13591	Taliaferro	1903	Wheeler	6500
Paulding	106350	Tattnall	23034	White	24015
Peach	24818	Taylor	8918	Whitfield	91842
Pickens	26810	Telfair	14685	Wilcox	8882
Pierce	16755	Terrell	10453	Wilkes	10613
Pike	15474	Thomas	42762	Wilkinson	9968
Polk	39564	Tift	38622	Worth	21873
Total					**8769252**

Table B.3: The oral cancer data by county in Georgia in 2004.

County	Observed Deaths	Expected Value	County	Observed Deaths	Expected Value	County	Observed Deaths	Expected Value
Appling	1	0.4436	Clinch	0	0.1716	Glynn	0	1.7618
Atkinson	0	0.1978	Cobb	10	16.1476	Gordon	1	1.2117
Bacon	0	0.2551	Coffee	3	0.9723	Grady	1	0.5995
Baker	0	0.1049	Colquitt	1	1.0805	Greene	0	0.3865
Baldwin	2	1.1162	Columbia	0	2.4836	Gwinnett	9	17.3028
Banks	0	0.3873	Cook	3	0.4013	Habersham	0	0.9624
Barrow	0	1.3930	Coweta	1	2.6018	Hall	7	3.9733
Bartow	2	2.1474	Crawford	0	0.3182	Hancock	0	0.2422
Ben Hill	1	0.4282	Crisp	1	0.5439	Haralson	2	0.6930
Berrien	2	0.4118	Dade	0	0.3948	Harris	1	0.6614
Bibb	6	3.8312	Dawson	0	0.4707	Hart	0	0.5770
Bleckley	0	0.2974	Decatur	0	0.7065	Heard	1	0.2788
Brantley	0	0.3837	De Kalb	11	16.6838	Henry	2	3.9382
Brooks	0	0.4041	Dodge	1	0.4815	Houston	5	3.0555
Bryan	1	0.6798	Dooly	2	0.2865	Irwin	2	0.2453
Bulloch	2	1.4899	Dougherty	4	2.3624	Jackson	2	1.2232
Burke	0	0.5725	Douglas	4	2.6472	Jasper	0	0.3177
Butts	0	0.5521	Early	0	0.2985	Jeff Davis	3	0.3165
Calhoun	0	0.1507	Echols	0	0.1013	Jefferson	1	0.4168
Camden	0	1.1137	Effingham	1	1.1027	Jenkins	2	0.2143
Candler	0	0.2517	Elbert	0	0.5162	Johnson	0	0.2367
Carroll	4	2.5080	Emanuel	1	0.5455	Jones	0	0.6477
Catoosa	3	1.4776	Evans	0	0.2777	Lamar	2	0.4052
Charlton	0	0.2641	Fannin	2	0.5336	Lanier	0	0.1843
Chatham	4	5.8891	Fayette	0	2.5019	Laurens	1	1.1532
Chattahoochee	0	0.3335	Floyd	2	2.3211	Lee	1	0.7386
Chattooga	0	0.6556	Forsyth	1	3.2558	Liberty	2	1.5246
Cherokee	0	4.3129	Franklin	0	0.5297	Lincoln	1	0.2073
Clarke	1	2.5666	Fulton	24	20.1087	Long	1	0.2698
Clay	0	0.0819	Gilmer	1	0.6606	Lowndes	0	2.3650
Clayton	0	6.5417	Glascock	1	0.0650	Lumpkin	0	0.5907

Table B.4: The oral cancer data by county in Georgia in 2004 continued.

County	Observed Deaths	Expected Value	County	Observed Deaths	Expected Value	County	Observed Deaths	Expected Value
McDuffie	2	0.5313	Pulaski	0	0.2429	Toombs	1	0.6611
McIntosh	0	0.2750	Putnam	1	0.4875	Towns	0	0.2502
Macon	0	0.3441	Quitman	0	0.0609	Treutlen	0	0.1739
Madison	0	0.6743	Rabun	0	0.3953	Troup	3	1.5111
Marion	0	0.1756	Randolph	0	0.1810	Turner	0	0.2321
Meriwether	0	0.5617	Richmond	10	4.8458	Twiggs	0	0.2580
Miller	0	0.1522	Rockdale	1	1.8967	Union	1	0.4841
Mitchell	0	0.5886	Schley	0	0.0998	Upson	3	0.6939
Monroe	1	0.5784	Screven	0	0.3787	Walker	3	1.5648
Montgomery	0	0.2215	Seminole	0	0.2288	Walton	1	1.7762
Morgan	1	0.4200	Spalding	1	1.5033	Ware	0	0.8793
Murray	0	1.0013	Stephens	1	0.6170	Warren	1	0.1544
Muscogee	4	4.5146	Stewart	1	0.1230	Washington	1	0.5200
Newton	0	2.0129	Sumter	2	0.8116	Wayne	2	0.6962
Oconee	1	0.7145	Talbot	0	0.1626	Webster	0	0.0574
Oglethorpe	0	0.3347	Taliaferro	0	0.0468	Wheeler	0	0.1627
Paulding	4	2.6156	Tattnall	3	0.5677	White	1	0.5826
Peach	1	0.6090	Taylor	3	0.2218	Whitfield	5	2.2088
Pickens	1	0.6857	Telfair	0	0.3188	Wilcox	0	0.2146
Pierce	0	0.41280	Terrell	0	0.2704	Wilkes	0	0.2613
Pike	0	0.3889	Thomas	2	1.0861	Wilkinson	3	0.2516
Polk	1	0.9942	Tift	0	0.9920	Worth	2	0.5434

B.2 South African Data

Table B.5: The South African data for acute pericarditis in 2014.

ID	Date of Birth	Age	Gender	Death Province	Smoker	Birth Province
1	09/07/1925	88	M	WC	Unavailable	EC
2	14/04/1947	66	F	EC	No	EC
3	16/12/1941	72	F	EC	Unspecified	EC
4	21/11/2008	5	F	NW	NA	NW
5	05/01/1928	85	F	EC	Unspecified	Unspecified
6	04/01/2007	7	M	LP	NA	LP
7	28/10/1963	50	M	LP	No	LP
8	27/01/1961	53	M	NW	Unspecified	Unspecified
9	28/12/1935	78	M	WC	Unspecified	GP
10	20/05/1964	50	F	NW	Yes	NW
11	16/11/1949	64	M	GP	Yes	FS
12	04/07/1976	38	M	FS	Yes	Foreign
13	11/03/1948	66	M	WC	No	WC
14	06/06/1985	29	M	GP	Yes	KZN
15	07/12/1989	24	F	NC	Yes	NC
16	23/11/1965	49	M	KZN	Unspecified	KZN

Appendix C: Additional Results

C.1 Additional results for Georgia model 1

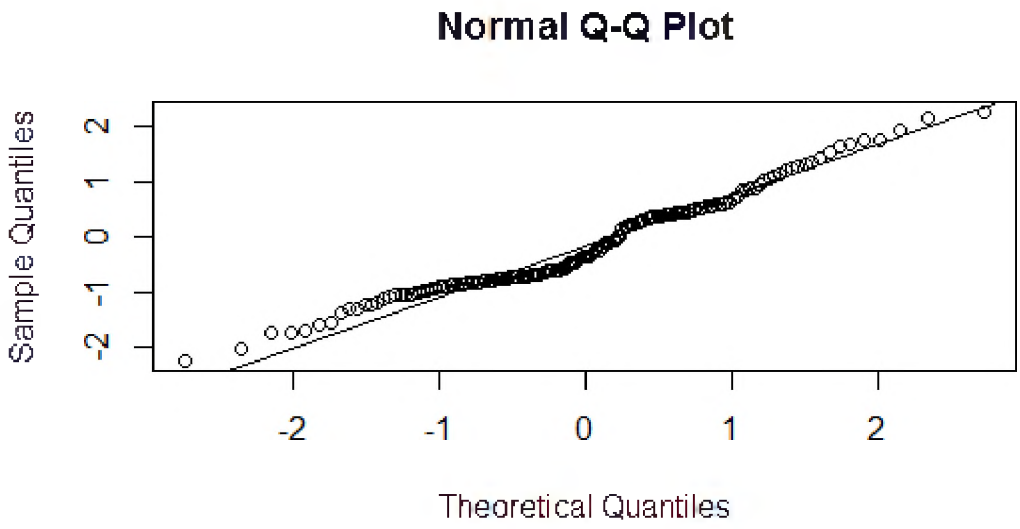


Figure C.1: Q-Q plot for model 1.

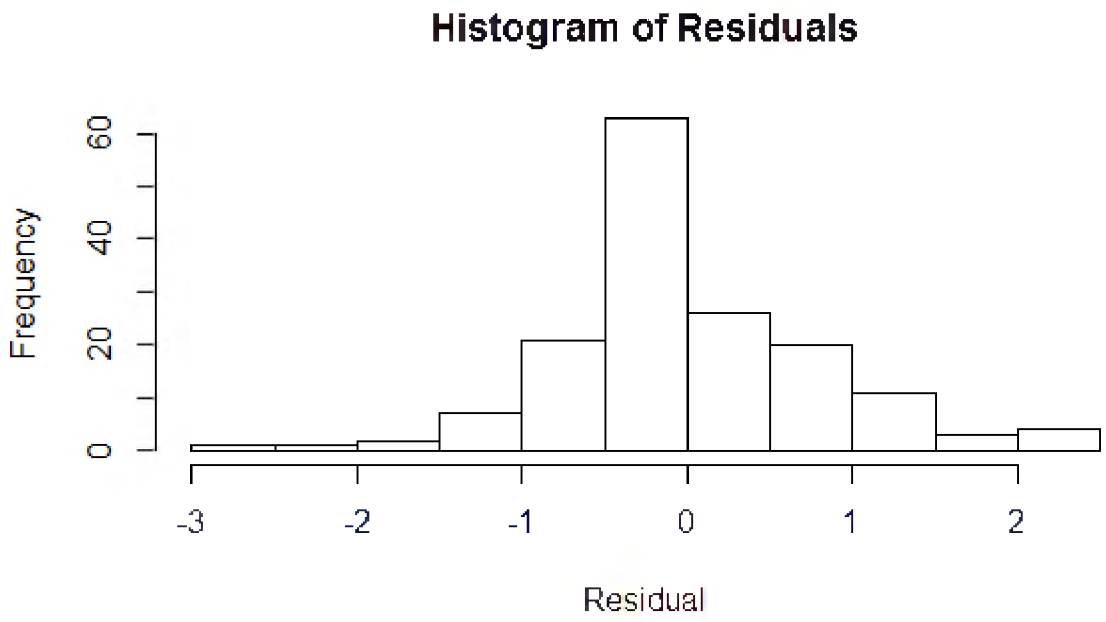


Figure C.2: Bayesian residual histogram for model 1.

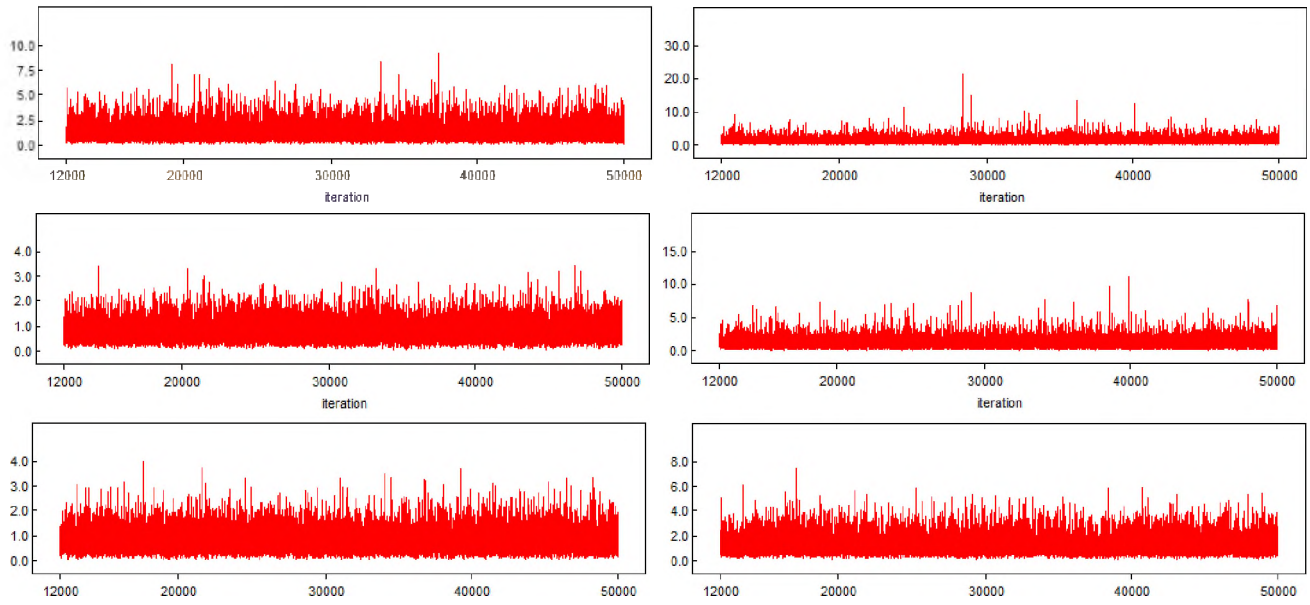


Figure C.3: Selected trace plots for the relative risk of model 1, from left to right: Bryan, Clinch, Coweta, Early, Glynn and Jackson.

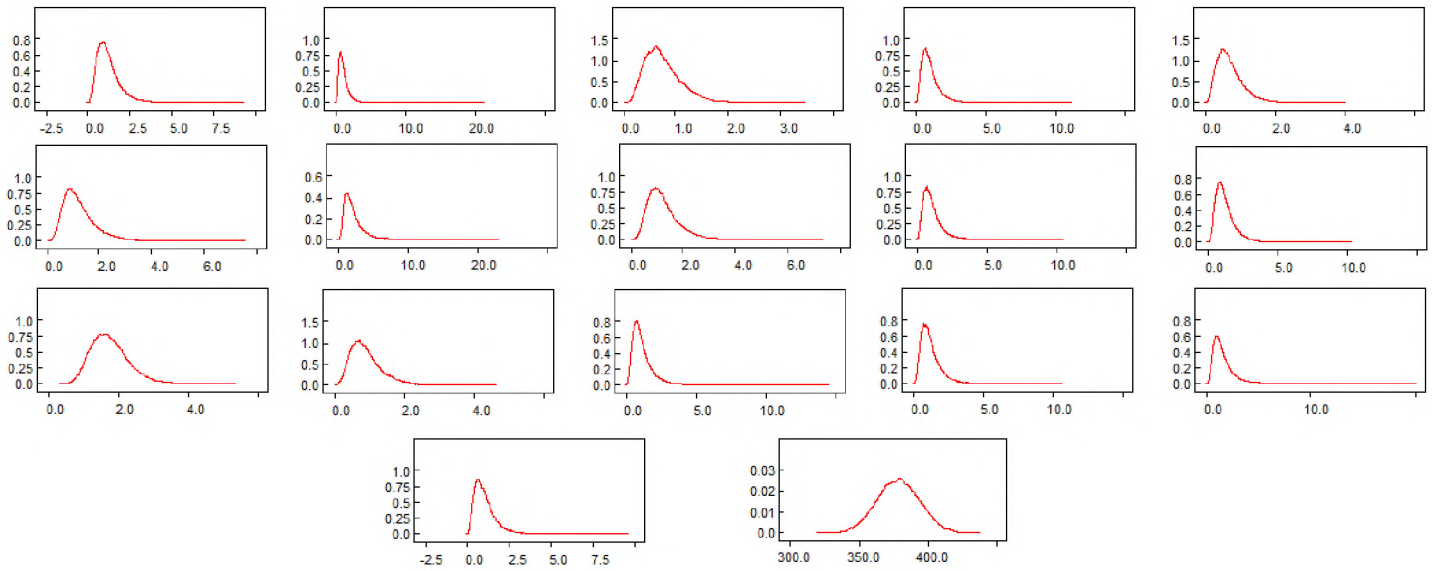


Figure C.4: Selected density kernels for the relative risk and deviance for model 1, from left to right: Bryan, Clinch, Coweta, Early, Glynn, Jackson, Jenkins, Liberty, McIntosh, Monroe, Richmond, Spalding, Taliaferro, Union, Warren, Wilkes and the deviance.

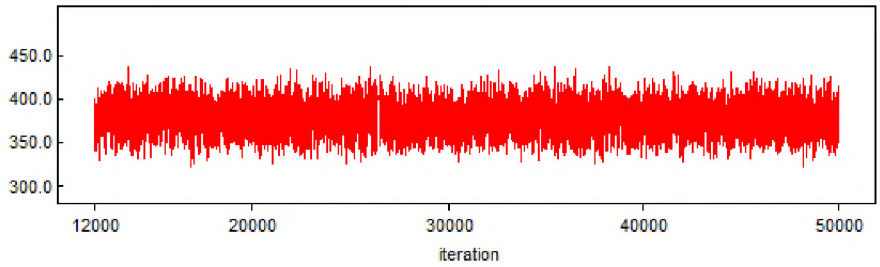


Figure C.5: Trace plot of the deviance for model 1.

Table C.1: Selected results for the deviance residuals in model 1.

County	Mean	Standard Deviation	MC Error	95% Confidence Interval	
				2.5%	97.5%
Bryan	0.2453	0.5050	0.003857	-0.8445	1.1520
Clinch	-0.5988	0.1893	0.001534	-1.0380	-0.2989
Coweta	-0.6758	0.5872	0.004773	-1.8730	0.4308
Early	-0.7609	0.2314	0.001978	-1.2900	-0.3843
Glynn	-1.5370	0.4130	0.004861	-2.4320	-0.8162
Jackson	0.4890	0.6278	0.007161	-0.8519	1.6360
Jenkins	1.7600	0.5560	0.006520	0.5031	2.6980
Liberty	0.1903	0.6449	0.005760	-1.1430	1.3840
McIntosh	-0.7420	0.2260	0.001932	-1.2610	-0.3744
Monroe	0.3738	0.4807	0.002922	-0.6701	1.2310
Richmond	0.5723	0.9028	0.006524	-1.2350	2.3000
Spalding	-0.2571	0.5469	0.005197	-1.3910	0.7574
Taliaferro	-0.3170	0.1011	6.438E-4	-0.5555	-0.1607
Union	0.5473	0.4756	0.004521	-0.4676	1.3940
Warren	1.2150	0.3963	0.003438	0.3325	1.9020
Wilkes	-0.7132	0.2150	0.001438	-1.2080	-0.3625

C.2 Additional results for Georgia model 2

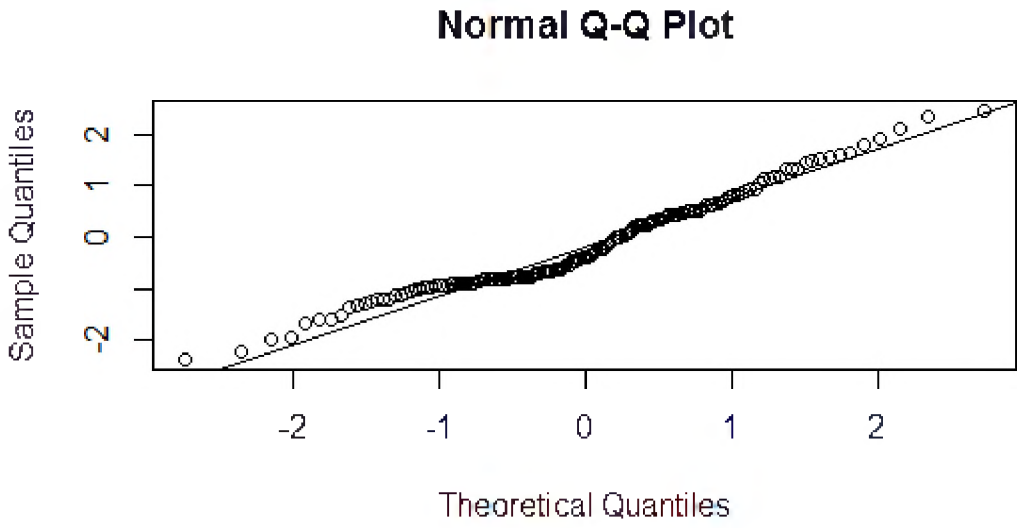


Figure C.6: Q-Q plot for model 2.

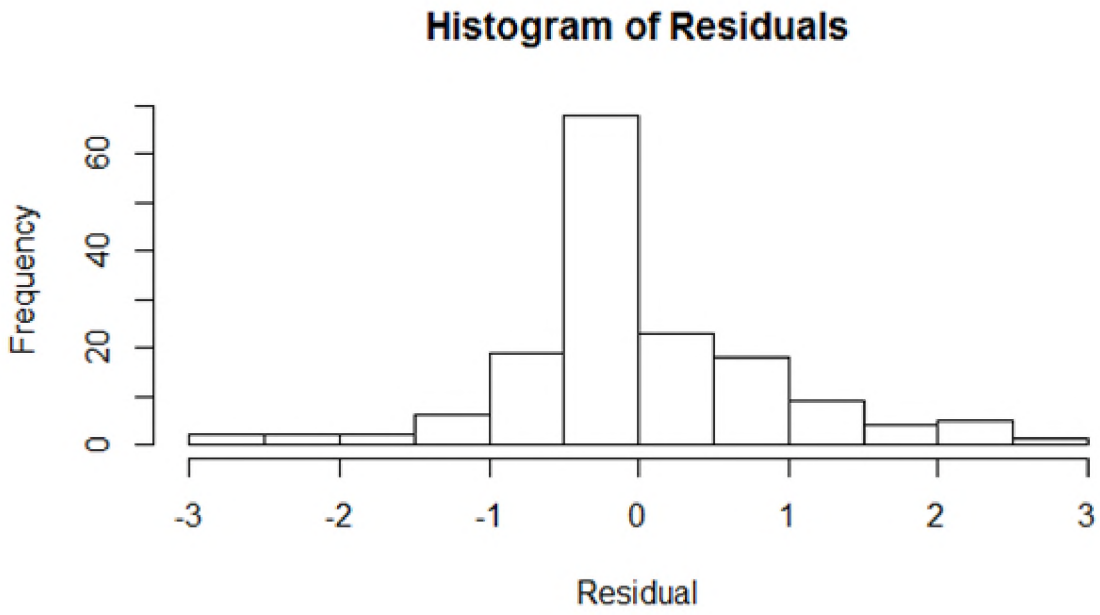


Figure C.7: Bayesian residual histogram for model 2.

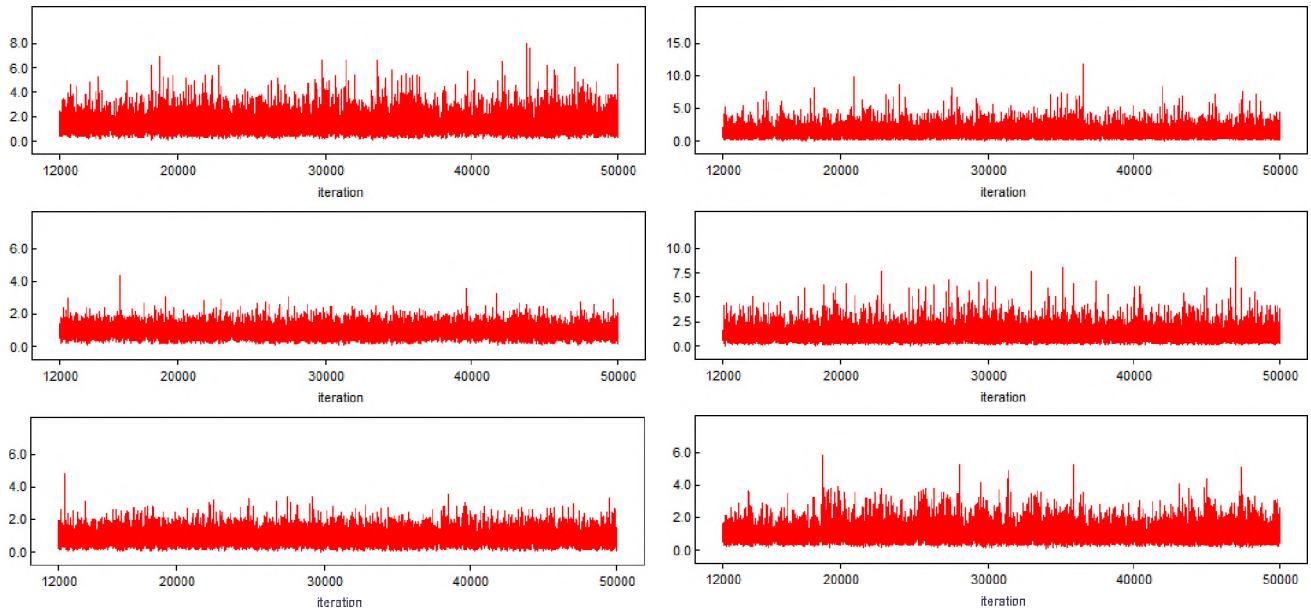


Figure C.8: Selected trace plots for the relative risk of model 2, from left to right: Bryan, Clinch, Coweta, Early, Glynn and Jackson.

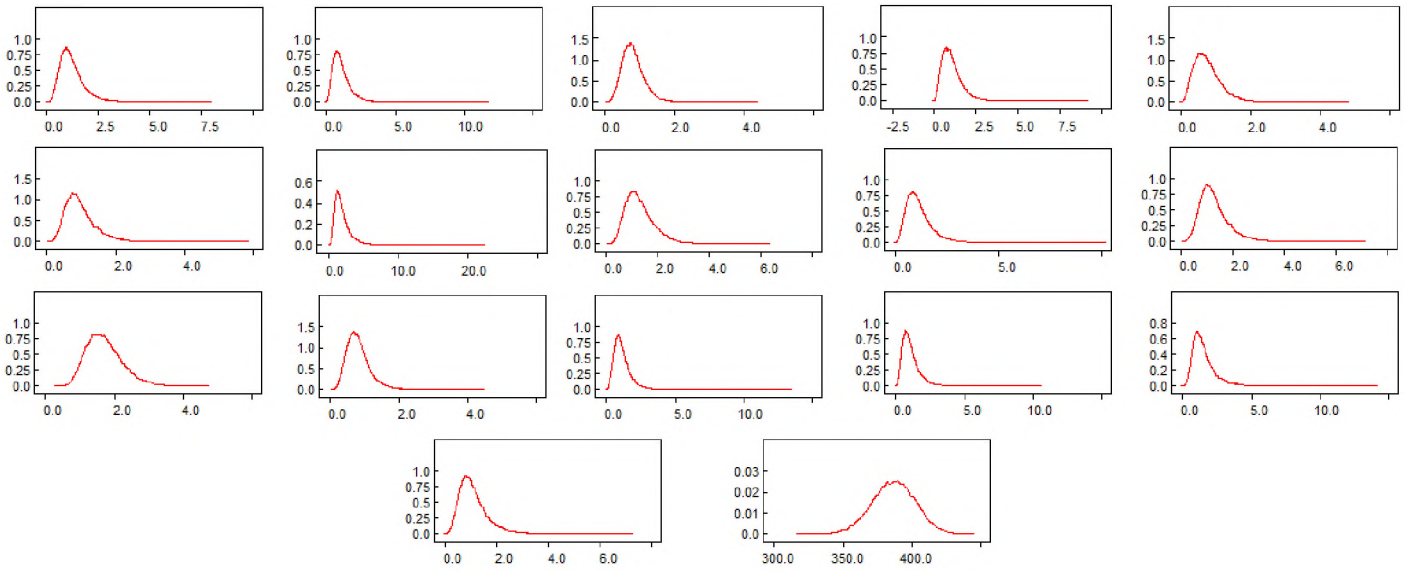


Figure C.9: Selected density kernels for the relative risk and deviance for model 2, from left to right: Bryan, Clinch, Coweta, Early, Glynn, Jackson, Jenkins, Liberty, McIntosh, Monroe, Richmond, Spalding, Taliaferro, Union, Warren, Wilkes and the deviance.

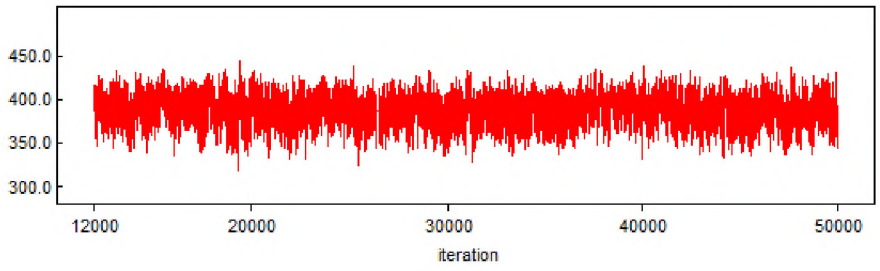


Figure C.10: Trace plot of the deviance for model 2.

Table C.2: Selected results for the deviance residuals in model 2.

County	Mean	Standard Deviation	MC Error	95% Confidence Interval	
				2.5%	97.5%
Bryan	0.2201	0.4372	0.003551	-0.7268	1.0220
Clinch	-0.6056	0.1741	0.001713	-1.0010	-0.3137
Coweta	-0.8346	0.5035	0.007783	-1.8420	0.1607
Early	-0.7724	0.2179	0.002897	-1.2540	-0.3982
Glynn	-1.6060	0.4038	0.006276	-2.4530	-0.8749
Jackson	0.7276	0.5189	0.008204	-0.3896	1.6800
Jenkins	1.7910	0.4888	0.006527	0.6775	2.6200
Liberty	0.0613	0.5927	0.005937	-1.1710	1.1680
McIntosh	-0.7778	0.2116	0.001932	-1.2610	-0.4234
Monroe	0.3273	0.3894	0.003213	-0.5122	1.0420
Richmond	0.6388	0.8797	0.008625	-1.1560	2.2920
Spalding	-0.1853	0.4305	0.004773	-1.0880	0.6341
Taliaferro	-0.3273	0.08351	7.218E-4	-0.5191	-0.1883
Union	0.6100	0.4329	0.003908	-0.3363	1.3810
Warren	1.1910	0.3314	0.003494	0.4445	1.7660
Wilkes	-0.7421	0.1840	0.001727	-1.1660	-0.4277

C.3 Additional results for Georgia model 3

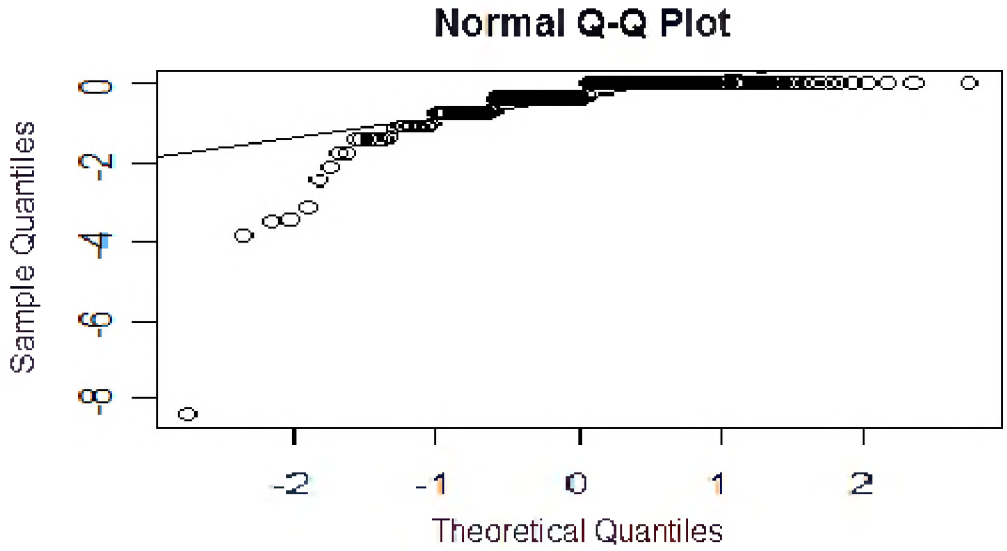


Figure C.11: Q-Q plot for model 3.

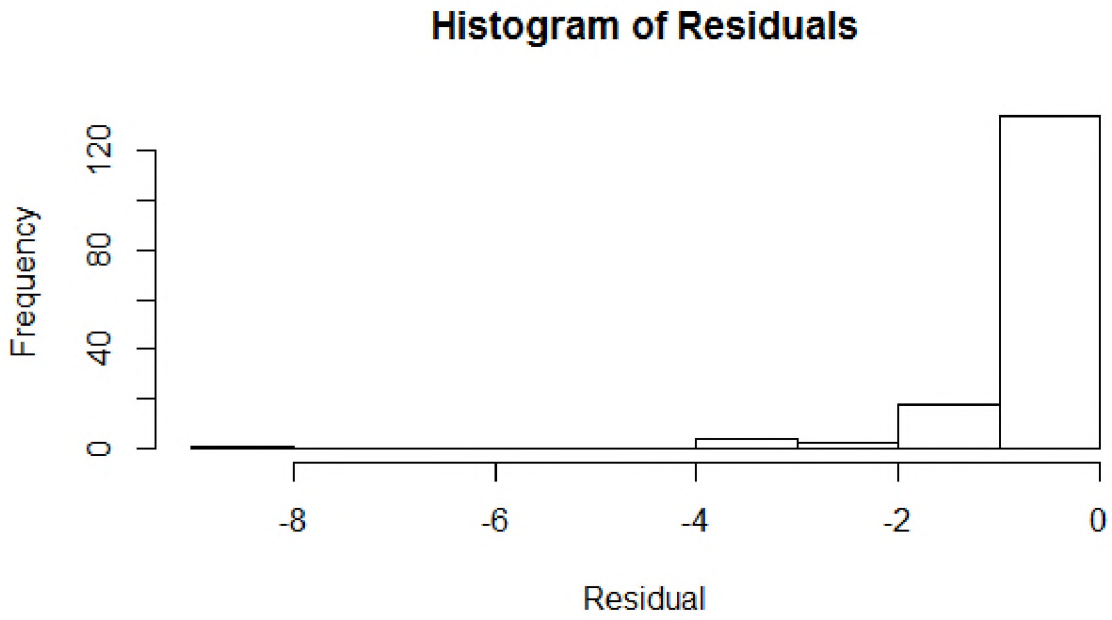


Figure C.12: Bayesian residual histogram for model 3.

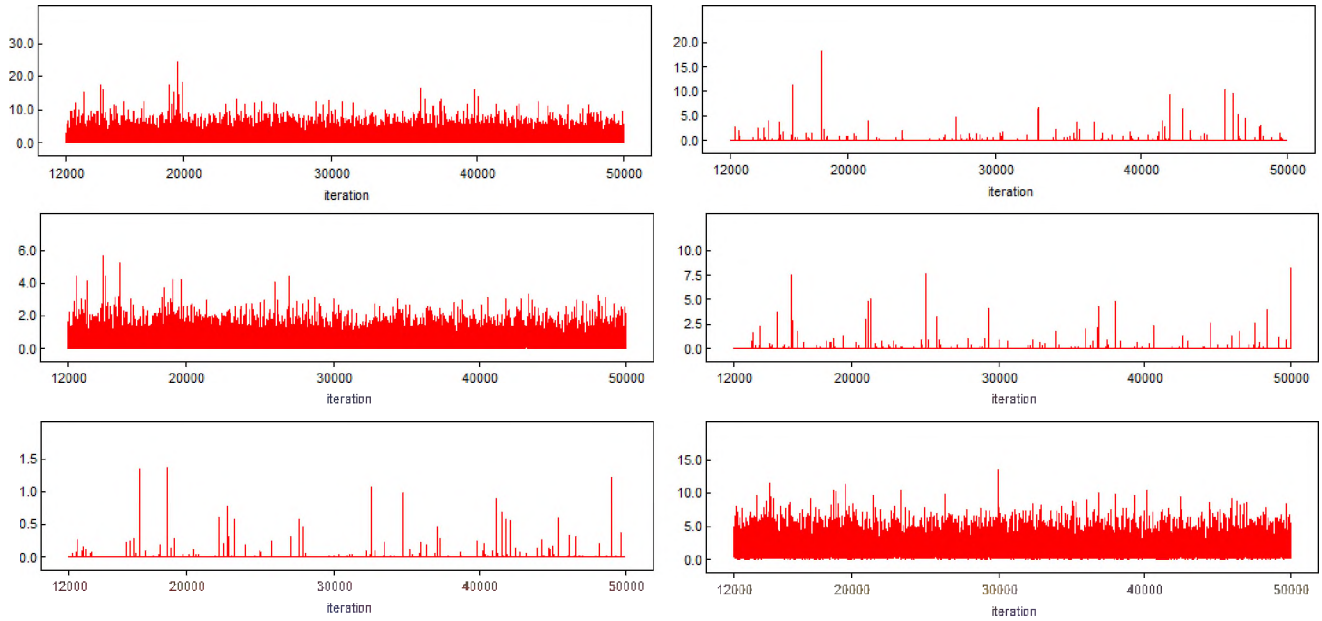


Figure C.13: Selected trace plots for the relative risk of model 3, from left to right: Bryan, Clinch, Coweta, Early, Glynn and Jackson.

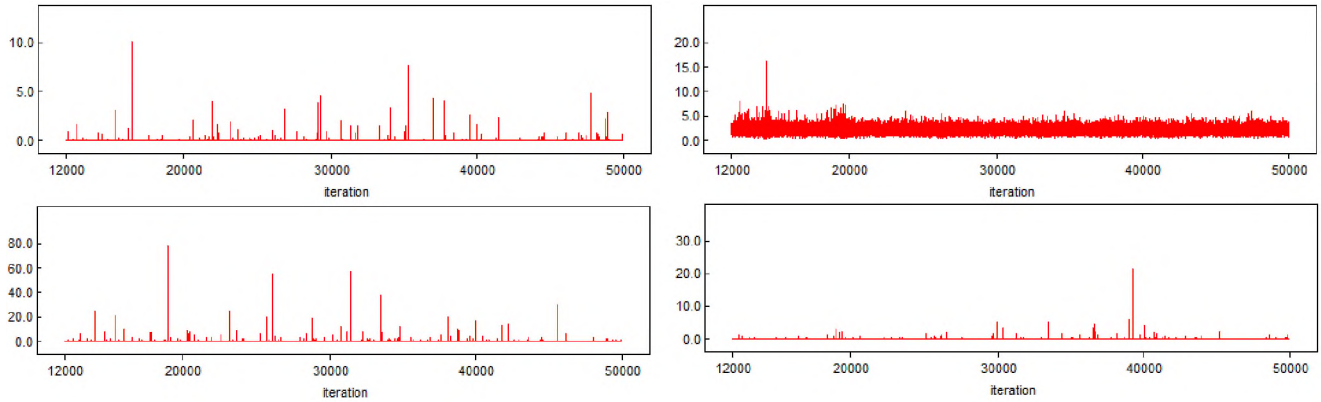


Figure C.14: Selected trace plots for the relative risk of model 3 continued, from left to right: McIntosh, Richmond, Taliaferro and Wilkes.

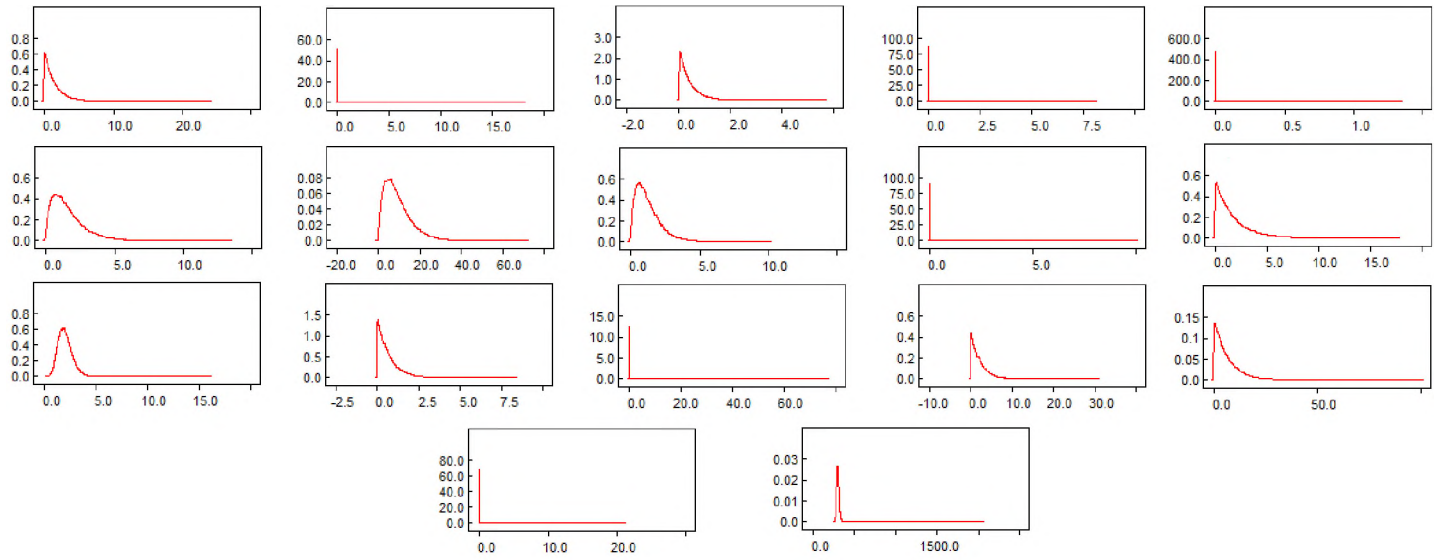


Figure C.15: Selected density kernels for the relative risk and deviance for model 3, from left to right: Bryan, Clinch, Coweta, Early, Glynn, Jackson, Jenkins, Liberty, McIntosh, Monroe, Richmond, Spalding, Taliaferro, Union, Warren, Wilkes and the deviance.

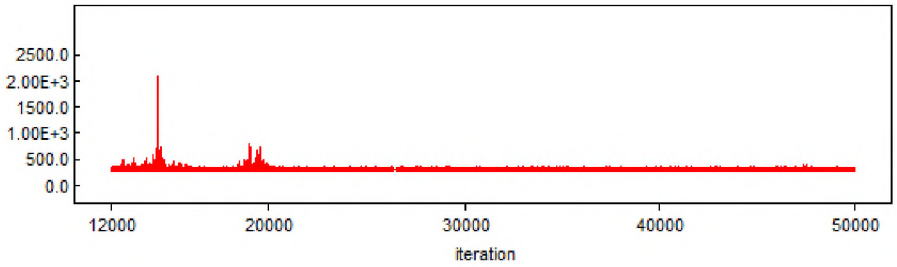


Figure C.16: Trace plot of the deviance for model 3.

Table C.3: Selected results for the deviance residuals in model 3.

County	Mean	Standard Deviation	MC Error	95% Confidence Interval	
				2.5%	97.5%
Bryan	0.3386	1.0220	0.005150	-1.686	2.299
Clinch	-0.0024	0.0437	2.309E-4	-1.701E-7	0
Coweta	0.3349	1.0270	0.004769	-1.684	2.329
Early	-0.0022	0.0440	2.155E-4	-4.787E-7	0
Glynn	-0.0023	0.0455	2.17E-4	-9.680E-8	0
Jackson	0.2333	1.0290	0.005160	-1.814	2.242
Jenkins	0.2347	1.0290	0.005519	-1.810	2.226
Liberty	0.2388	1.0260	0.004748	-1.805	2.233
McIntosh	-0.0020	0.0400	2.014E-4	-3.625E-7	0
Monroe	0.3382	1.0350	0.005563	-1.713	2.367
Richmond	0.0971	1.0640	0.005385	-2.011	2.131
Spalding	0.3426	1.0280	0.004985	-1.697	2.331
Taliaferro	-0.0022	0.0425	1.999E-4	-2.141E-7	0
Union	0.3490	1.0250	0.005376	-1.675	2.340
Warren	0.3292	1.0220	0.004963	-1.696	2.321
Wilkes	-0.0021	0.0389	2.022E-4	-1.213E-7	0

C.4 Additional Results for the ICAR model

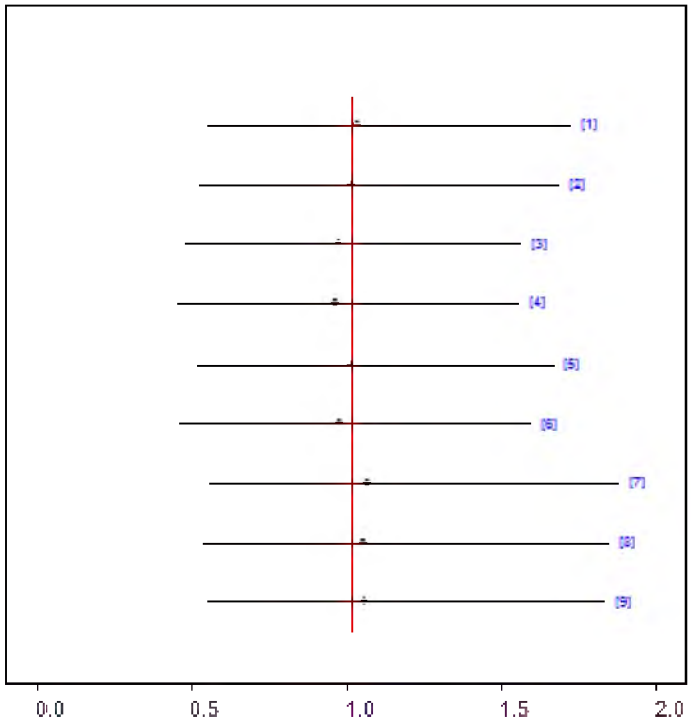


Figure C.17: Caterpillar plot for the relative risk for the South African data for the ICAR model.

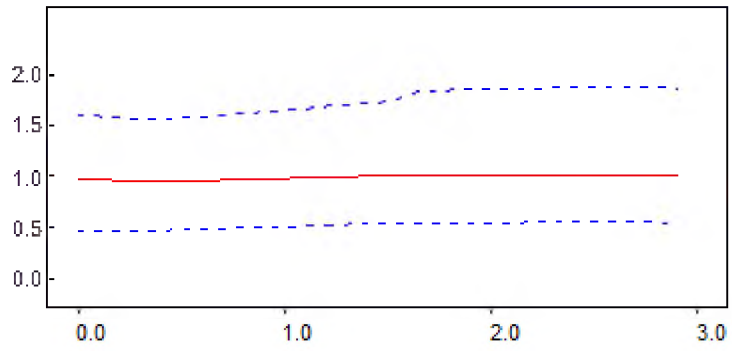


Figure C.18: SMR and relative risk model fit for South African data for the ICAR model.

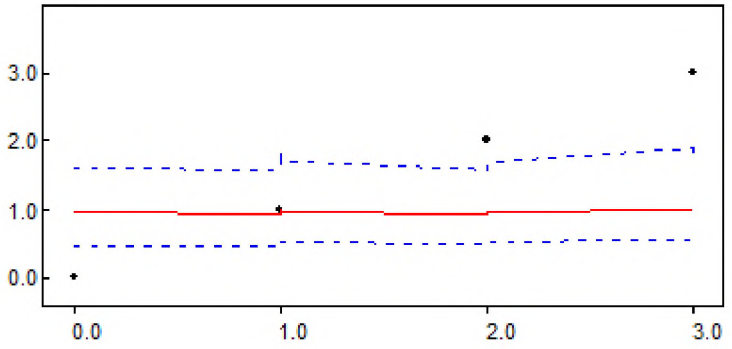


Figure C.19: Relative risk and observed value model fit for the South African data for the ICAR model.

C.5 Additional Results for the PCAR model

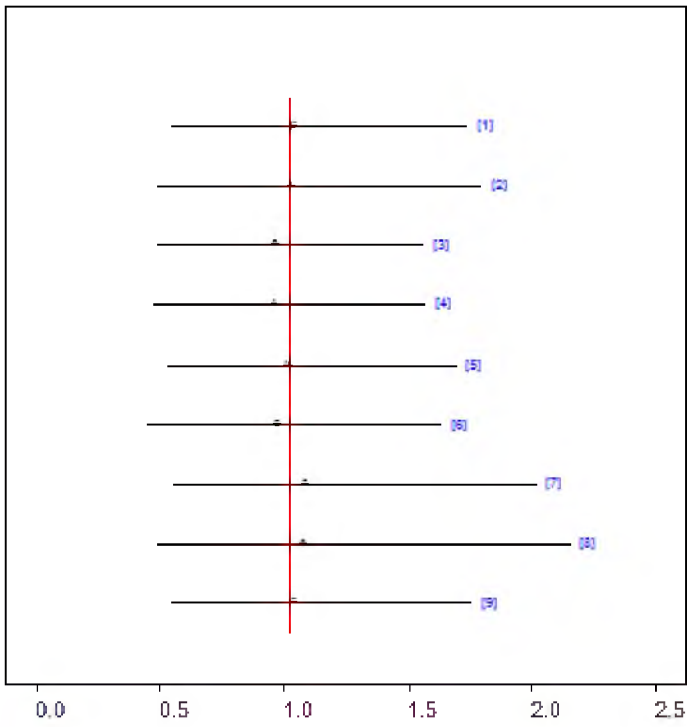


Figure C.20: Caterpillar plot for the relative risk for the South African data for the PCAR model.

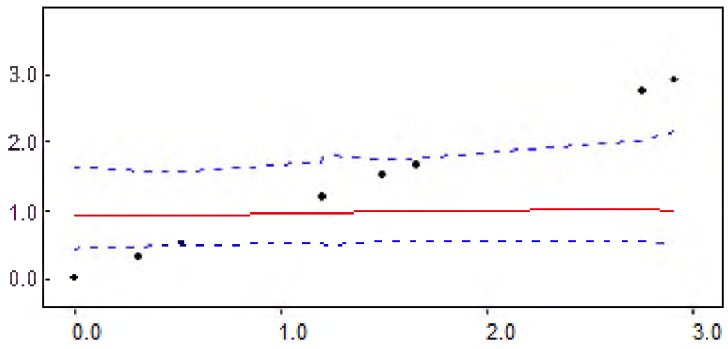


Figure C.21: SMR and the relative risk model fit for South African data for the PCAR model.

Appendix D: Code

```
The code for the neighbouring regions in South Africa.
library(spdep)
library(maptools)
x<-readShapeSpatial("C:/Users/Downloads/AppData/Local/Temp/ZAF_adm_shp
+/ZAF_adm1.shp")
x<-x[x$NAME_1,]
x$value<-c(5,4,30,12,2,0,3,1,11)
par(mai=c(0,0,0,0))
xy<-coordinates(x)
w<-poly2nb(x,row.names=x$ID_1)
summary(w)
plot(x,col='gray',border='blue',lwd=2)
plot(w,xy,col='red',lwd=2,add=TRUE)

The code for the neighbouring regions in Georgia.
library(spdep)
library(maptools)
x<-readShapeSpatial("C:/Users/Richard/Documents/US/USG.shp")
x<-x[x$NAME_2,]
x$value<-c(1,0,0,0,2,0,0,2,1,2,6,0,0,0,1,2,0,0,0,0,0,4,3,0,4,0,0,0,1,0,
0,0,10,3,1,0,3,1,0,1,0,0,0,11,1,2,4,4,0,0,1,0,1,0,2,0,2,1,0,24,1,1,0,1,
1,0,9,0,7,0,2,1,0,1,2,5,2,2,0,3,1,2,0,0,2,0,1,1,2,1,1,0,0,2,0,0,0,0,0,0,
0,1,0,1,0,4,0,1,0,4,1,1,0,0,1,0,1,0,0,0,10,1,0,0,0,1,1,1,2,0,0,3,3,0,0,
2,0,1,0,0,3,0,0,1,3,3,1,0,1,1,2,0,0,1,5,0,0,3,2)
par(mai=c(0,0,0,0))
xy<-coordinates(x)
w<-poly2nb(x,row.names=x$ID_2)
summary(w)
plot(x,col='gray',border='blue',lwd=2)
```

```
plot(w,xy,col='red',lwd=2,add=TRUE)
```

The code for the line graph comparing standard populations.

```
age<-c('0-4','05-9','10-14','15-19','20-24','25-29','30-34','35-39','40-44',
'45-49','50-54','55-59','60-64','65-69','70-74','75-79','80-84','85+')
age<-as.factor(age)
world<-c(8.86,8.69,8.6,8.47,8.22,7.93,7.61,7.15,6.59,6.04,5.37,4.55,3.72,
2.96,2.21,1.52,0.91,0.63)
africa<-c(16.25,14.74,13.13,10.81,8.41,6.9,5.73,4.76,4.03,3.44,2.86,2.45,
2.03,1.65,1.22,0.81,0.44,0.35)
asia<-c(12.19,12.02,11.9,10.78,8.72,7.26,6.51,5.93,5.31,4.56,3.78,3.19,2.68,
2.07,1.47,0.89,0.46,0.28)
segi<-c(12,10,9,9,8,8,6,6,6,6,5,4,4,3,2,1,0.5,0.5)
scand<-c(8,7,7,7,7,7,7,7,7,7,7,6,5,4,3,2,1,1)
plot(age,africa,main="Comparison of standard populations",xlab="Age Group",
ylab="Percentage of Population",col="red")
lines(age,world,col="blue")
lines(age,asia,col='black')
lines(age,segi,col="green")
lines(age,scand,col="orange")
leg<-c("Africa","World","Asia","Segi","Scandinavian")
col<-c('Red','Blue','Black','Green','Orange')
legend("topright",leg,col=col,lwd=3)
```

The code for the Poisson density function figure.

```
x<-seq(0,20,1)
plot(dpois(x,1),main="Poisson probability mass function",xlab="Time",
ylab="",type="1",col="black",lwd=3)
lines(dpois(x,4),col="red",lwd=3)
lines(dpois(x,5),col="green",lwd=3)
lines(dpois(x,10),col="orange",lwd=3)
colour<-c("black","red","green","orange")
labels=c(expression(paste(lambda,"=1"),paste(lambda,"=2"),paste(lambda,"=5")
,paste(lambda,"=10")))
legend("topright",inset=0.00,labels,col=colour,lwd=3)
```

The code for the Poisson cumulative distribution function.

```
x<-seq(0,20,1)
plot(ppois(x,1),main="Poisson cumulative distribution function",xlab="Time"
,ylab="",type="1",col="black",lwd=3)
lines(ppois(x,4),col="red",lwd=3)
lines(ppois(x,5),col="green",lwd=3)
lines(ppois(x,10),col="orange",lwd=3)
colour<-c("black","red","green","orange")
labels=c(expression(paste(lambda,"=1"),paste(lambda,"=2"),paste(lambda,"=5")
,paste(lambda,"=10")))
legend("bottomright",inset=0.0,labels,col=colour,lwd=3)
```

The code for the calculation of parameters in the conjugate prior distribution for the Georgia data.

```
y=c(1,0,0,0,2,0,0,2,1,2,6,0,0,0,1,2,0,0,0,0,0,4,3,0,4,0,0,0,1,0,0,0,10,3,1,0
,3,1,0,1,0,0,0,11,1,2,4,4,0,0,1,0,1,0,2,0,2,1,0,24,1,1,0,1,1,0,9,0,7,0,2,1,0
,1,2,5,2,2,0,3,1,2,0,0,2,0,1,1,2,1,1,0,0,2,0,0,0,0,0,0,0,1,0,1,0,4,0,1,0,4,1
,1,0,0,1,0,1,0,0,0,10,1,0,0,0,1,1,1,2,0,0,3,3,0,0,2,0,1,0,0,3,0,0,1,3,3,1,0,
1,1,2,0,0,1,5,0,0,3,2)
a=mean(y[1:46])
b=var(y[1:46])
b=a/b
a=b*a
```

The code for the conjugate prior simulation with the South African parameters.

```
theta<-rep(1:9,1)
n<-10000
count2<-rep(0,length(theta))
length<-rep(0,n)
sd2<-rep(0,length(theta))
lengthbar2<-rep(0,length(theta))
theta2<-matrix(c(rep(0,n)),ncol=n)
for (j in 1:length(theta)){
for (i in 1:n){
y<-rpois(n,theta[j])
theta2<-rgamma(n,shape=y+13.78723,rate=1.765957)
sorted<-sort(theta2)
```

```
intervallower<-sorted[n*0.025]
intervalhigher<-sorted[n*0.975]
length[i]<-intervalhigher-intervallower
a<-n*0.025
b<-n*0.975
if (theta2[a]>intervallower){
if (theta2[b]<intervalhigher){
count2[j]<-count2[j]+1 }}}
lengthbar2[j]<-(1/n)*sum(length)
sd2[j]<-sqrt((1/(n-1))*(sum((length^2))-(n*lengthbar2[j]^2)))}
```

The code for the conjugate prior simulation with the Georgia parameters.

```
theta<-rep(1:9,1)
n<-10000
count6<-rep(0,length(theta))
length<-rep(0,n)
sd6<-rep(0,length(theta))
lengthbar6<-rep(0,length(theta))
theta6<-matrix(c(rep(0,n)),ncol=n)
for (j in 1:length(theta)){
for (i in 1:n){
y<-rpois(n,theta[j])
theta6<-rgamma(n,shape=y+0.3139,rate=1.2329)
sorted<-sort(theta6)
intervallower<-sorted[n*0.025]
intervalhigher<-sorted[n*0.975]
length[i]<-intervalhigher-intervallower
a<-n*0.025
b<-n*0.975
if (theta6[a]>intervallower){
if (theta6[b]<intervalhigher){
count6[j]<-count6[j]+1}}}
lengthbar6[j]<-(1/n)*sum(length)
sd6[j]<-sqrt((1/(n-1))*(sum((length^2))-(n*lengthbar6[j]^2)))}
```

The code for coverage rate plot of the conjugate prior distributions.

```
plot(theta,count4/n,ylim=range(c(count4/n,count6/n)),type="o",lwd=2,
col="black",xlab=expression(paste(theta)),ylab="Coverage Rate",
main="The coverage rate of conjugate prior distributions")
lines(theta,count2/n,col="red",lwd=2,type="o")
lines(theta,count3/n,col="green",lwd=2,type='o')
lines(theta,count5/n,col="orange",lwd=2,type="o")
lines(theta,count6/n,col="purple",lwd=2,type="o")
colour<-c("black","red","green","orange","purple")
lab1=expression(paste(alpha,"=0.5"," ",beta,"=6"))
lab2=expression(paste(alpha,"=8"," ",beta,"=0.25"))
lab3=expression(paste(alpha,"=3"," ",beta,"=5"))
lab4=expression(paste(alpha,"=13.78723"," ",beta,"=0.765957"))
lab5=expression(paste(alpha,"=0.3139"," ",beta,"=0.2329"))
labels=c(lab1,lab2,lab3,lab4,lab5)
legend("right",inset=0,labels,col=colour,lwd=2)
```

The code for standard deviation of the conjugate prior distributions.

```
plot(theta,sd4,ylim=range(c(sd4,sd6)),type="o",lwd=2,col="black",
xlab=expression(paste(theta)),ylab="Standard Deviation",main="The
standard deviation of conjugate prior distributions")
lines(theta,sd2,col="red",lwd=2,type="o")
lines(theta,sd3,col="green",lwd=2,type='o')
lines(theta,sd4,col="orange",lwd=2,type="o")
lines(theta,sd5,col="purple",lwd=2,type="o")
colour<-c("black","red","green","orange","purple")
lab1=expression(paste(alpha,"=0.5"," ",beta,"=6"))
lab2=expression(paste(alpha,"=8"," ",beta,"=0.25"))
lab3=expression(paste(alpha,"=3"," ",beta,"=5"))
lab4=expression(paste(alpha,"=13.78723"," ",beta,"=0.765957"))
lab5=expression(paste(alpha,"=0.3139"," ",beta,"=0.2329"))
labels=c(lab1,lab2,lab3,lab4,lab5)
legend("right",inset=0,labels,col=colour,lwd=2)
```

The code for mean interval length of the conjugate prior distributions.

```
plot(theta,lengthbar4,ylim=range(c(lengthbar4,lengthbar6)),type="o",lwd=2,
col="black",xlab=expression(paste(theta)),ylab="Average Interval Length"
```

```
,main="The average interval length of conjugate prior distributions")
lines(theta,lengthbar2,col="red",lwd=2,type="o")
lines(theta,lengthbar3,col="green",lwd=2,type='o')
lines(theta,lengthbar5,col="orange",lwd=2,type="o")
lines(theta,lengthbar6,col="purple",lwd=2,type="o")
colour<-c("black","red","green","orange","purple")
lab1=expression(paste(alpha,"=0.5"," ",beta,"=6"))
lab2=expression(paste(alpha,"=8"," ",beta,"=0.25"))
lab3=expression(paste(alpha,"=3"," ",beta,"=5"))
lab4=expression(paste(alpha,"=13.78723"," ",beta,"=0.765957"))
lab5=expression(paste(alpha,"=0.3139"," ",beta,"=0.2329"))
labels=c(lab1,lab2,lab3,lab4,lab5)
legend("right",inset=0,labels,col=colour,lwd=2)
```

The code for the simulation using the Jeffreys' Prior.

```
theta<-rep(1:9,1)
n<-10000
count7<-rep(0,length(theta))
length<-rep(0,n)
sd7<-rep(0,length(theta))
lengthbar7<-rep(0,length(theta))
theta7<-matreix(c(rep(0,n)),ncol=n)
for (j in 1:length(theta)){
for (i in 1:n){
y<-rpois(1,theta[j])
theta7<-rgamma(n=n,shape=y+0.5,rate=1)
sorted<-sort(theta7)
intervallower<-sorted[n*0.025]
intervalhigher<-sorted[n*0.975]
length[i]<-intervalhigher-intervallower
a<-n*0.025
b<-n*0.975
if (theta7[a]>intervallower){
if (theta7[b]<intervalhigher){
count7[j]<-count7[j]+1}}}
lengthbar7[j]<-(1/n)*sum(length)
sd7[j]<-sqrt((1/(n-1))*(sum(length^2)-((lengthbar7[j]^2)*n)))}
```

The code for the simulation using the probability matching prior.

```
theta<-rep(1:9,1)
n<-10000
count8<-rep(0,length(theta))
length<-rep(0,n)
sd8<-rep(0,length(theta))
lengthbar8<-rep(0,length(theta))
theta8<-matrix(c(rep(0,n)),ncol=n)
for (j in 1:length(theta)){
for (i in 1:n){
y<-rpois(1,theta[j])
theta8<-rgamma(n=n,shape=y-(1/2)+1,rate=1)
sorted<-sort(theta8)
intervallower<-sorted[n*0.025]
intervalhigher<-sorted[n*0.975]
length[i]<-intervalhigher-intervallower
a<-n*0.025
b<-n*0.975
if (theta8[a]>intervallower){
if (theta8[b]<intervalhigher){
count8[j]<-count8[j]+1}}}
lengthbar8[j]<-(1/n)*sum(length)
sd8[j]<-sqrt((1/(n-1))*(sum(length^2)-((lengthbar8[j]^2)*n)))}
```

The code for the simulation using the divergence prior.

```
theta<-rep(1:9,1)
n<-10000
count9<-rep(0,length(theta))
length<-rep(0,n)
sd9<-rep(0,length(theta))
lengthbar9<-rep(0,length(theta))
theta9<-matrix(c(rep(0,n)),ncol=n)
for (j in 1:length(theta)){
for (i in 1:n){
y<-rpois(1,theta[j])
theta9<-rgamma(n=n,shape=y+0.75,rate=1)
```

```
sorted<-sort(theta9)
intervallower<-sorted[n*0.025]
intervalhigher<-sorted[n*0.975]
length[i]<-intervalhigher-intervallower
a<-n*0.025
b<-n*0.975
if (theta9[a]>intervallower){
if (theta9[b]<intervalhigher){
count9[j]<-count9[j]+1}}}
lengthbar9[j]<-(1/n)*sum(length)
sd9[j]<-sqrt((1/(n-1))*(sum(length^2)-((lengthbar9[j]^2)*n)))}
```

The code for coverage rate line graph of all the prior distributions.

```
plot(theta,count5/n,ylim=range(c(count6/n,count9/n)),type="o",lwd=2
,col="orange",xlab=expression(paste(theta)),ylab="Coverage Rate",
main="The coverage rate for the prior distributions")
lines(theta,count6/n,col="purple",lwd=2,type="o")
lines(theta,count7/n,col="green",lwd=2,type='o')
lines(theta,count8/n,col="black",lwd=2,type="o")
lines(theta,count9/n,col="red",lwd=2,type="o")
colour<-c("orange","purple","green","black","red")
labels=c("Conjugate South Africa","Conjugate Georgia","Jeffreys'",
"Probability Matching","Divergence")
legend("bottomright",inset=0,labels,col=colour,lwd=2)
```

The code for average interval length line graph of all the prior distributions.

```
plot(theta,lengthbar5,ylim=range(c(lengthbar6,lengthbar9)),type="o",lwd=2
,col="orange",xlab=expression(paste(theta)),ylab="Average Interval
Length",main="The average interval length for the prior distributions")
lines(theta,lengthbar6,col="purple",lwd=2,type="o")
lines(theta,lengthbar7,col="green",lwd=2,type='o')
lines(theta,lengthbar8,col="black",lwd=2,type="o")
lines(theta,lengthbar9,col="red",lwd=2,type="o")
colour<-c("orange","purple","green","black","red")
labels=c("Conjugate South Africa","Conjugate Georgia","Jeffreys'",
```

```
"Probability Matching","Divergence")
legend("bottomright",inset=0,labels,col=colour,lwd=2)

The code for standard deviation line graph of all the prior distributions.
plot(theta,sd5,ylim=range(c(sd6,sd7)),type="o",lwd=2,
col="orange",xlab=expression(paste(theta)),ylab="Standard Deviation",
main="The standard deviation for the prior distributions")
lines(theta,sd6,col="purple",lwd=2,type="o")
lines(theta,sd7,col="green",lwd=2,type='o')
lines(theta,sd8,col="black",lwd=2,type="o")
lines(theta,sd9,col="red",lwd=2,type="o")
colour<-c("orange","purple","green","black","red")
labels=c("Conjugate South Africa","Conjugate Georgia","Jeffreys'",
"Probability Matching","Divergence")
legend("bottomright",inset=0,labels,col=colour,lwd=2)

The code for model 1.
model{
for(i in 1:N) {
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
smr[i]<-(y[i]+eps2)/(e[i]+eps2)
theta[i]<-exp(B0+v[i]+S[i])
v[i]~dnorm(0,phi)
PP[i]<-step(theta[i]-1+eps)
ypred[i]~dpois(mu[i])
PPL[i]<-pow(ypred[i]-y[i],2)
PPL2[i]<-abs(ypred[i]-y[i])
res[i]<-(y[i]-mu[i])
Lexp[i]<-log(e[i])
Ds[i]<-2*((y[i]*log(y[i]/mu[i]))-(y[i]-mu[i]))
sign[i]<-2*step(y[i]-mu[i])-1
dev.res[i]<-sign[i]*sqrt(Ds[i])}
S[1:N]~car.normal(adj[],weights[],num[],omega)
for(k in 1:sumNumNeigh){
weights[k]<-1}
B0~dflat()
```

```
eps<-1.0E-6
eps2~dnorm(0,1000)
phi~dgamma(0.3139,0.2329)
mape<-mean(PPL2[])
mspe<-mean(PPL[])
omega~dgamma(0.1,0.0001)
sdS<-sd(S[])
sdV<-sd(v[])
sdE<-sd(Lexp[])
sum<-sdS*sdS+sdV*sdV+sdE*sdE
pS<-sdS*sdS/sum
pV<-sdV*sdV/sum
pE<-sdE*sdE/sum}
```

The code for the first part of the calculation of Moran's I in RStudio.

```
num=c(6,5,5,6,5,6,6,7,5,7,6,6,6,4,5,7,5,6,6,3,5,6,2,3,2,5,3,7,5,4,5,4,5,
8,6,3,5,7,5,6,1,7,5,5,6,6,6,4,5,3,4,5,10,5,5,4,4,5,4,10,6,4,4,9,3,6,7,6,
9,7,3,4,3,3,6,7,5,6,6,7,8,4,6,7,5,5,7,5,5,4,4,5,6,7,4,6,7,6,7,4,7,7,5,6,
4,3,6,7,7,6,5,5,5,4,4,5,6,3,2,6,4,5,4,4,3,8,3,4,8,7,6,8,7,6,6,4,6,7,4,6,
4,6,6,4,7,5,6,7,6,6,7,6,6,5,4,7,6,7,7)
csum<-rep(0,160)
csum[1]<-0
x<-rep(0,159)
g<-rep(0,159)
for(i in 2:(N+1)){
csum[i]<-sum(num[1:(i-1)])}
for(i in 1:159){
x[i]<-csum[i]+1
g[i]<-csum[i+1]}
```

The code for the second part of the calculation of Moran's I in WinBUGS and calculated for model 1.

```
model {
for(i in 1:N){
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
theta[i]<-exp(B0+v[i]+S[i])
```

```
v[i]~dnorm(0,phi)
res[i]<-(y[i]-mu[i])/sqrt(mu[i])
estar[i]<-sum(We[x[i]:g[i]])
de[i]<-res[i]-mean(res[])
d.estar[i]<-estar[i]-mean(estar[])
dt[i]<-de[i]*d.estar[i]
db[i]<-pow(d.estar[i],2)}
S[1:N]~car.normal(adj[],weights[],num[],omega)
for(k in 1:sumNumNeigh){
weights[k]<-1
We[k]<-res[adj[k]]}
Moran<-sum(dt[])/sum(db[])
B0~dflat()
phi~dgamma(0.3139,0.2329)
omega~dgamma(0.1,0.0001)}
list(N=159,y=c(1,0,0,0,2,0,0,2,1,2,6,0,0,0,1,2,0,0,0,0,0,4,3,0,4,0,0,0,1,0,
0,0,10,3,1,0,3,1,0,1,0,0,0,11,1,2,4,4,0,0,1,0,1,0,2,0,2,1,0,24,1,1,0,1,1,0,
9,0,7,0,2,1,0,1,2,5,2,2,0,3,1,2,0,0,2,0,1,1,2,1,1,0,0,2,0,0,0,0,0,0,0,1,0,1,
0,4,0,1,0,4,1,1,0,0,1,0,1,0,0,0,10,1,0,0,0,1,1,1,2,0,0,3,3,0,0,2,0,1,0,0,3
,0,0,1,3,3,1,0,1,1,2,0,0,1,5,0,0,3,2),e=c(0.4436,0.1978,0.2551,0.1049,
1.1162,0.3873,1.393,2.1474,0.4282,0.4118,3.8312,0.2974,0.3837,0.4041,0.6798,
1.4899,0.5725,0.5521,0.1507,1.1137,0.2517,2.508,1.4776,0.2641,5.8891,0.3335,
0.6556,4.3129,2.5666,0.0819,6.5417,0.1716,16.1476,0.9723,1.0805,2.4836,
0.4013,2.6018,0.3182,0.5439,0.3948,0.4707,0.7065,16.6838,0.4815,0.2865,
2.3624,2.6472,0.2985,0.1013,1.1027,0.5162,0.5455,0.2777,0.5336,2.5019,2.3211
,3.2558,0.5297,20.1087,0.6606,0.065,1.7618,1.2117,0.5995,0.3865,17.3028,
0.9624,3.9733,0.2422,0.693,0.6614,0.577,0.2788,3.9382,3.0555,0.2453,1.2232,
0.3177,0.3165,0.4168,0.2143,0.2367,0.6477,0.4052,0.1843,1.1532,0.7386,1.5246
,0.2073,0.2698,2.365,0.5907,0.5313,0.275,0.3441,0.6743,0.1756,0.5617,0.1522,
0.5886,0.5784,0.2215,0.42,1.0013,4.5146,2.0129,0.7145,0.3347,2.6156,0.609,
0.6857,0.4128,0.3889,0.9942,0.2429,0.4875,0.0609,0.3953,0.181,4.8458,1.8967,
0.0998,0.3787,0.2288,1.5033,0.617,0.123,0.8116,0.1626,0.0468,0.5677,0.2218,
0.3188,0.2704,1.0861,0.992,0.6611,0.2502,0.1739,1.5111,0.2321,0.258,0.4841,
0.6939,1.5648,1.7762,0.8793,0.1544,0.52,0.6962,0.0574,0.1627,0.5826,2.2088,
0.2146,0.2613,0.2516,0.5434),num=c(6,5,5,6,5,6,6,7,5,7,6,6,6,4,5,7,5,6,6,3,5
,6,2,3,2,5,3,7,5,4,5,4,5,8,6,3,5,7,5,6,1,7,5,5,6,6,6,4,5,3,4,5,10,5,5,4,4,5,
4,10,6,4,4,9,3,6,7,6,9,7,3,4,3,3,6,7,5,6,6,7,8,4,6,7,5,5,7,5,5,4,4,5,6,7,4,6
```

```
,7,6,7,4,7,7,5,6,4,3,6,7,7,6,5,5,5,4,4,5,6,3,2,6,4,5,4,4,3,8,3,4,8,7,6,8,7,6
,6,4,6,7,4,6,4,6,6,4,7,5,6,7,6,6,7,6,6,5,4,7,6,7,7),adj=c(151,138,132,113,80
,3,148,86,34,32,10,148,113,80,34,1,101,100,49,47,43,19,158,150,117,84,70,127
,97,78,69,68,59,147,108,78,69,67,29,115,112,110,64,57,33,28,156,142,134,77,
34,137,92,86,77,37,34,2,143,111,102,84,76,39,158,143,116,87,76,45,151,148,
113,63,24,20,136,92,37,35,89,54,51,25,16,124,82,54,53,51,21,15,124,121,82,81
,53,126,107,102,85,79,75,135,120,49,47,30,4,63,24,13,138,132,54,53,16,110,74
,71,60,48,38,155,146,148,20,13,51,15,152,130,128,106,98,146,64,57,112,64,60,
58,42,33,8,109,108,97,78,7,120,118,49,19,126,75,60,56,44,148,86,50,2,110,60,
48,28,8,148,134,80,77,10,9,3,2,159,137,136,101,37,14,121,94,90,137,92,35,14,
10,141,126,99,74,60,56,22,145,133,111,102,11,159,156,142,129,88,46,146,112,
93,69,61,58,55,28,125,101,100,65,4,122,75,67,60,31,156,153,134,116,87,12,156
,129,116,96,76,40,159,135,101,88,19,4,110,60,33,22,125,100,30,19,4,92,86,32,
124,25,16,15,157,109,97,90,73,140,138,132,103,83,82,81,21,17,16,132,89,21,16
,15,144,105,93,61,42,126,60,38,31,115,64,27,8,69,67,60,42,28,127,97,73,6,67,
58,56,48,44,38,33,31,28,22,112,105,93,64,55,42,150,149,81,70,151,95,20,13,
155,146,112,105,61,57,28,27,8,136,101,43,131,117,109,108,104,70,147,122,69,
60,58,44,7,154,139,127,119,69,6,154,93,78,68,67,58,42,7,6,150,149,131,117,
66,62,5,115,110,22,141,130,106,99,97,59,52,141,38,22,126,122,107,44,31,18,
143,116,111,96,46,12,11,142,137,34,10,9,108,97,69,29,7,6,117,107,104,102,84,
18,153,138,134,103,34,3,1,150,149,121,94,83,62,53,17,124,53,17,16,158,150,
140,87,81,53,158,143,117,102,79,11,5,145,126,114,102,18,92,50,32,10,2,158,
153,143,140,83,45,12,159,135,129,47,40,132,95,91,54,15,157,94,52,36,151,132,
95,89,86,50,37,14,10,154,144,69,61,55,42,157,149,131,121,90,81,36,151,91,89,
63,133,129,123,111,76,46,109,78,73,59,52,29,6,152,133,130,129,123,26,145,141
,130,126,114,72,38,125,49,43,4,159,136,65,47,43,35,4,145,85,84,79,39,18,11,
153,140,138,80,53,147,117,108,107,79,66,155,64,61,55,130,72,26,147,122,104,
79,75,18,147,109,104,78,66,29,7,157,131,108,97,66,52,29,115,71,48,33,22,8,
133,96,76,39,11,64,61,42,28,8,151,148,13,3,1,145,126,99,85,110,71,57,8,156,
76,46,45,12,104,84,79,70,66,5,128,120,30,139,68,152,135,128,118,30,19,94,81
,36,17,147,107,75,67,44,133,129,98,96,82,51,17,16,100,49,43,114,99,85,75,56
,38,31,18,68,59,6,152,120,118,26,152,135,123,98,96,88,46,40,145,133,106,99,
98,72,26,157,149,109,94,70,66,151,138,91,89,54,53,21,1,145,130,123,111,98,
96,39,156,153,80,45,34,9,152,129,120,88,47,19,101,65,35,14,159,142,77,37,35
,10,140,132,103,80,53,21,1,154,144,119,68,153,138,103,87,83,53,99,74,72,38,
159,156,137,77,40,9,158,87,84,76,12,11,154,139,93,55,133,130,114,102,99,85
,39,155,64,41,27,23,122,108,107,104,67,7,113,34,32,24,13,3,2,157,131,94,81
```

```
,70,62,158,83,81,70,62,5,132,113,95,91,63,13,1,135,129,128,120,98,26,140,
134,103,87,80,45,144,139,93,69,68,146,105,64,23,142, 134,116,46,45,40,9,149
,131,109,94,90,52,150,143,87,84,83,12,5,142,137,101,88,47,40,35),
x=c(1,7,12,17,23,28,34,40,47,52,59,65,71,77,81,86,93,98,104,110,113,118,124
,126,129,131,136,139,146,151,155,160,164,169,177,183,186,191,198,203,209,210
,217,222,227,233,239,245,249,254,257,261,266,276,281,286,290,294,299,303,313
,319,323,327,336,339,345,352,358,367,374,377,381,384,387,393,400,405,411,417
,424,432,436,442,449,454,459,466,471,476,480,484,489,495,502,506,512,519,525
,532,536,543,550,555,561,565,568,574,581,588,594,599,604,609,613,617,622,628
,631,633,639,643,648,652,656,659,667,670,674,682,689,695,703,710,716,722,726
,732,739,743,749,753,759,765,769,776,781,787,794,800,806,813,819,825,830,834
,841,847,854),g=c(6,11,16,22,27,33,39,46,51,58,64,70,76,80,85,92,97,103,109,
112,117,123,125,128,130,135,138,145,150,154,159,163,168,176,182,185,190,197,
202,208,209,216,221,226,232,238,244,248,253,256,260,265,275,280,285,289,293,
298,302,312,318,322,326,335,338,344,351,357,366,373,376,380,383,386,392,399,
404,410,416,423,431,435,441,448,453,458,465,470,475,479,483,488,494,501,505,
511,518,524,531,535,542,549,554,560,564,567,573,580,587,593,598,603,608,612,
616,621,627,630,632,638,642,647,651,655,658,666,669,673,681,688,694,702,709,
715,721,725,731,738,742,748,752,758,764,768,775, 780,786,793,799,805,812,818
,824,829,833,840,846,853,860),sumNumNeigh=860))
```

QQ Plot model 1

```
dev.res<-c(0.5202,-0.6437,-0.7259,-0.4703,0.4572,-0.821,-1.377,0.01233,
0.5329,1.266,0.3806,-0.7815,-0.8475,-0.884,0.2453,0.2114,-1.033,-0.9709,
-0.5624,-1.291,-0.7248,0.4651,0.6116,-0.7145,-0.2837,-0.8094,-1.046,-2.005,
-0.6366,-0.4191,-2.228,-0.5988,-0.2528,0.9457,-0.09747,-1.732,1.747,-0.6758
,-0.801,0.3724,-0.8488,-0.896,-1.099,-0.149,0.4634,1.541,0.4537,0.4304,
-0.7609,-0.4647,-0.09601,-0.9412,0.3759,-0.7563,1.144,-1.682,-0.06153,
-0.8034,-0.9363,0.2262,0.3344,1.682,-1.537,-0.1083,0.3367,-0.8428,-0.3574,
-1.2,0.5973,-0.6977,0.8951,0.2863,-0.9731,0.883,-0.444,0.3959,1.652,0.489,
-0.7714,1.941,0.5702,1.76,-0.7056,-1.073,1.325,-0.6192,-0.1613,0.1688,0.1903
,1.06,0.8585,-1.728,-0.9946,1.09,-0.742,-0.8384,-1.039,-0.6109,-0.9939,
-0.5579,-1.035,0.3738,-0.6863,0.6405,-1.236,-0.07562,-1.574,0.285,-0.7825,
0.435,0.2998,0.3129,-0.8887,-0.8522,0.03006,-0.7164,0.5073,-0.3666,-0.823,
-0.6154,0.5723,-0.3893,-0.473,-0.8639,-0.6717,-0.2571,0.4035,1.345,0.7118,
-0.5839,-0.317,1.432,2.266,-0.8108,-0.7419,0.492,-1.292,0.2347,-0.6745,
-0.6106,0.6317,-0.7024,-0.7324,0.5473,1.287,0.6005,-0.3605,-1.209,1.215,
```

```
0.4159,0.8591,-0.3578,-0.595,0.4297,0.7449,-0.676,-0.7132,2.157,1.045)
qqnorm(dev.res)
qqline(dev.res)

Code for model 2.
model{
for(i in 1:N){
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
smr[i]<-(y[i]+eps2)/(e[i]+eps2)
theta[i]<-exp(B0+v[i]+S[i])
v[i]~dnorm(0,phi)
PP[i]<-step(theta[i]-1+eps)
ypred[i]~dpois(mu[i])
PPL[i]<-pow(ypred[i]-y[i],2)
PPL2[i]<-abs(ypred[i]-y[i])
res[i]<-(y[i]-mu[i])
Lexp[i]<-log(e[i])
Ds[i]<-2*((y[i]*log(y[i]/mu[i]))-(y[i]-mu[i]))
sign[i]<-2*step(y[i]-mu[i])-1
dev.res[i]<-sign[i]*sqrt(Ds[i])}
S[1:N]~car.normal(adj[],weights[],num[],omega)
for(k in 1:sumNumNeigh){
weights[k]<-1}
B0~dflat()
eps<-1.0E-6
eps2~dnorm(0,1000)
phi~dgamma(0.5,0.0001)
mape<-mean(PPL2[])
mspe<-mean(PPL[])
omega~dgamma(0.1,0.0001)
sdS<-sd(S[])
sdV<-sd(v[])
sdE<-sd(Lexp[])
sum<-sdS*sdS+sdV*sdV+sdE*sdE
pS<-sdS*sdS/sum
pV<-sdV*sdV/sum
```

```
pE<-sdE*sdE/sum }

QQ Plot model 2
dev.res<-c(0.419,-0.6846,-0.8028,-0.475,0.3343,-0.7834,-1.381,0.05458,0.3664
,1.314,0.263,-0.8911,-0.8341,-0.9591,0.2201,0.07184,-1.196,-0.965,-0.5833,
-1.257,-0.8119,0.5082,0.4625,-0.6797,-0.345,-0.8817,-1.076,-2.216,-0.6345,
-0.4324,-2.37,-0.6056,-0.3944,0.9504,-0.2736,-1.949,1.936,-0.8346,-0.9166,
0.2067,-0.8816,-0.8855,-1.146,-0.008875,0.3359,1.532,0.4436,0.5127,-0.772,
-0.4645,-0.1621,-0.9306,0.2403,-0.8281,1.339,-1.685,-0.05562,-0.8463,-0.8816
,0.6723,0.3973,1.622,-1.606,-0.12,0.3203,-0.863,-0.45,-1.202,1.141,-0.7694,
0.8973,0.247,-0.9123,0.878,-0.3322,0.2196,1.632,0.7276,-0.7767,2.102,0.4476,
1.791,-0.8069,-1.208,1.477,-0.6376,-0.4149,0.008625,0.06131,1.127,0.8064,
-1.985,-1.023,1.16,-0.7778,-0.9857,-0.9987,-0.6837,-1.058,-0.5514,-1.126,
0.3273,-0.7841,0.77,-1.331,-0.2249,-1.601,0.427,-0.7634,0.5153,0.1181,0.4231
,-0.942,-0.8949,-0.01653,-0.8196,0.5069,-0.3906,-0.7783,-0.6559,0.6388,
-0.2359,-0.5429,-0.9535,-0.6641,-0.1853,0.5599,1.342,0.6179,-0.6319,-0.3273,
1.582,2.458,-0.9428,-0.8095,0.4564,-1.53,0.05552,-0.6518,-0.6867,0.6892,
-0.8086,-0.8237,0.61,1.434,0.6474,-0.2521,-1.299,1.191,0.2525,0.9504,-0.3862
,-0.6743,0.5221,0.8356,-0.7776,-0.7421,2.363,1.06)
qqnorm(dev.res)
qqline(dev.res)

The code for the calculation of Moran's I for model 2.
model {
for(i in 1:N){
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
theta[i]<-exp(B0+v[i]+S[i])
v[i]~dnorm(0,phi)
res[i]<-(y[i]-mu[i])/sqrt(mu[i])
estar[i]<-sum(We[x[i]:g[i]])
de[i]<-res[i]-mean(res[])
d.estar[i]<-estar[i]-mean(estar[])
dt[i]<-de[i]*d.estar[i]
db[i]<-pow(d.estar[i],2)}
S[1:N]~car.normal(adj[],weights[],num[],omega)
for(k in 1:sumNumNeigh){
```

```
weights[k]<-1
We[k]<-res[adj[k]]}
Moran<-sum(dt[])/sum(db[])
B0~dflat()
phi~dgamma(0.5,0.0001)
omega~dgamma(0.1,0.0001)}

Code for model 3.
model{
for(i in 1:N){
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
smr[i]<-(y[i]+eps2)/(e[i]+eps2)
theta[i]<-exp(B0+v[i]+S[i])
v[i]~dnorm(0,1.0E-6)
PP[i]<-step(theta[i]-1+eps)
ypred[i]~dpois(mu[i])
PPL[i]<-pow(ypred[i]-y[i],2)
PPL2[i]<-abs(ypred[i]-y[i])
res[i]<-(y[i]-mu[i])
Lexp[i]<-log(e[i])
Ds[i]<-2*((y[i]*log(y[i]/mu[i]))-(y[i]-mu[i]))
sign[i]<-2*step(y[i]-mu[i])-1
dev.res[i]<-sign[i]*sqrt(Ds[i])}
S[1:N]~car.normal(adj[],weights[],num[],omega)
for(k in 1:sumNumNeigh){
weights[k]<-1}
B0~dflat()
eps<-1.0E-6
eps2~dnorm(0,1000)
mape<-mean(PPL2[])
mspe<-mean(PPL[])
omega~dgamma(0.005,0.005)
sdS<-sd(S[])
sdV<-sd(v[])
sdE<-sd(Lexp[])
sum<-sdS*sdS+sdV*sdV+sdE*sdE
```

```
pS<-sdS*sdS/sum
pV<-sdV*sdV/sum
pE<-sdE*sdE/sum}
```

The code for the calculation of Moran's I for model 3.

```
model {
for(i in 1:N){
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
theta[i]<-exp(B0+v[i]+S[i])
v[i]~dnorm(0,1.0E-6)
res[i]<-(y[i]-mu[i])/sqrt(mu[i])
estar[i]<-sum(We[x[i]:g[i]])
de[i]<-res[i]-mean(res[])
d.estar[i]<-estar[i]-mean(estar[])
dt[i]<-de[i]*d.estar[i]
db[i]<-pow(d.estar[i],2)}
S[1:N]~car.normal(adj[],weights[],num[],omega)
for(k in 1:sumNumNeigh){
weights[k]<-1
We[k]<-res[adj[k]]}
Moran<-sum(dt[])/sum(db[])
B0~dflat()
omega~dgamma(0.005,0.005)}
```

Code for caluclation of expected values of the South African Data.

```
population<-c(6786880,2786757,12914817,10694434,5630464,4229323,3676274
,1166680,6116324)
obs<-c(3,1,2,1,2,0,3,1,3) #ec,FS,GP,kzn,LP,mp,nw,nc,wc
total<-54001953
totalobs<-16
rate<-totalobs/total
e<-rate*population
```

Code for South African ICAR model.

```
model{
for(i in 1:N){
```

```
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
smr[i]<-(y[i]+eps2)/(e[i]+eps2)
theta[i]<-exp(B0+v[i]+S[i])
v[i]~dnorm(0,phi)
PP[i]<-step(theta[i]-1+eps)
ypred[i]~dpois(mu[i])
PPL[i]<-pow(ypred[i]-y[i],2)
PPL2[i]<-abs(ypred[i]-y[i])
res[i]<-(y[i]-mu[i])/sqrt(mu[i])
Lexp[i]<-log(e[i])
Ds[i]<-2*((y[i]*log(y[i]/mu[i]))-(y[i]-mu[i]))
sign[i]<-2*step(y[i]-mu[i])-1
dev.res[i]<-sign[i]*sqrt(Ds[i])}}
S[1:N]~car.normal(adj[],weights[],num[],omega)
for(k in 1:sumNumNeigh){
weights[k]<-1}
B0~dflat()
eps<-1.0E-6
eps2~dnorm(0,1000)
mape<-mean(PPL2[])
mspe<-mean(PPL[])
omega~dgamma(0.5,0.0005)
sdS<-sd(S[])
sdV<-sd(v[])
sdE<-sd(Lexp[])
phi~dgamma(0.5,0.0005)
sum<-sdS*sdS+sdV*sdV+sdE*sdE
pS<-sdS*sdS/sum
pV<-sdV*sdV/sum
pE<-sdE*sdE/sum
QR60<-ranked(theta[],8)/ranked(theta[],2)}
list(N=9,y=c(3,1,2,1,2,0,3,1,3),e=c(2.0108547,0.8256759,3.8264741,3.1686066,
1.6682253,1.2530874,1.0892270,0.3456705,1.8121786),num=c(4,6,4,3,3,4,4,4,2),
adj=c(9,8,4,2,8,7,6,4,3,1,7,6,5,2,6,2,1,7,6,3,5,4,3,2,8,5,3,2,9,7,2,1,8,1),
sumNumNeigh=34))
```

Code for QQ plot ICAR.

```
dev.res<-c(0.6448,0.2039,-0.9309,-1.324,0.2725,-1.543,1.473,0.8958,0.7723,
-0.002066)
qqnorm(dev.res,lwd=2)
qqline(dev.res)
```

Residual vs fitted value plot code for the ICAR model.

```
res<-c(0.9258,0.165,-1.71,-2.037,0.3113,-1.215,1.838,0.637,1.082)
fitted<-c(2.851,1.805,3.808,3.442,2.569,2.177,2.127,1.188,2.729)
plot(fitted,res,main="Residuals versus Fitted values",ylab="Residual",
xlab="Fitted Value",lwd=4,type='p',cex=.5)
```

Code for South African PCAR model.

```
model{
for (i in 1:N){
M[i]<-1/e[i]}
csum[1]<-0
for(i in 2:(N+1)){
csum[i]<-sum(num[1:(i-1)]) }
for(k in 1:sumNumNeigh)
{for(i in 1:N){
pick[k,i]<-step(k-csum[i]-epsilon)*step(csum[i+1]-k)}
W[k]<-sqrt(e[adj[k]]/inprod(e[],pick[k,]))}
epsilon<-0.0001
for (i in 1:N){
y[i]~dpois(mu[i])
mu[i]<-e[i]*theta[i]
theta[i]<-exp(S[i]+B0+v[i])
v[i]~dnorm(0,phi)
mu1[i]<-alpha
Lexp[i]<-log(e[i])
smr[i]<-(y[i]+eps2)/(e[i]+eps2)
PP[i]<-step(theta[i]-1+eps)
ypred[i]~dpois(mu[i])
PPL[i]<-pow(ypred[i]-y[i],2)
PPL2[i]<-abs(ypred[i]-y[i])
res[i]<-(y[i]-mu[i])
```

```
Ds[i]<-2*((y[i]*log(y[i]/mu[i]))-(y[i]-mu[i]))
sign[i]<-2*step(y[i]-mu[i])-1
dev.res[i]<-sign[i]*sqrt(Ds[i])}
S[1:N]~car.proper(mu1[],W[],adj[],num[],M[],omega,rho)
phi~dgamma(0.5,0.0005)
alpha~dnorm(0,0.0001)
omega~dgamma(0.5,0.0005)
dev.sat<-sum(Ds[])
eps<-1.0E-6
eps2~dnorm(0,1000)
mape<-mean(PPL2[])
mspe<-mean(PPL[])
B0~dflat()
rho.min<-min.bound(W[],adj[],num[],M[])
rho.max<-max.bound(W[],adj[],num[],M[])
rho~dunif(rho.min, rho.max)
sdS<-sd(S[])
sdV<-sd(v[])
sdE<-sd(Lexp[])
sum<-sdS*sdS+sdV*sdV+sdE*sdE
pS<-sdS*sdS/sum
pV<-sdV*sdV/sum
pE<-sdE*sdE/sum
QR60<-ranked(theta[],8)/ranked(theta[],2)}
list(N=9,y=c(3,1,2,1,2,0,3,1,3),e=c(2.0108547,0.8256759,3.8264741,3.1686066,
1.6682253,1.2530874,1.0892270,0.3456705,1.8121786),num=c(4,6,4,3,3,4,4,4,2),
adj=c(9,8,4,2,8,7,6,4, 3,1,7,6,5,2,6,2,1,7,6,3,5,4,3,2,8,5,3,2,9,7,2,1,8,1),
sumNumNeigh=34))
```

Code for QQplot PCAR

```
dev.res<-c(0.654,0.2029,-0.9363,-1.336,0.2737,-1.547,1.463,0.8957,0.8055)
qqnorm(dev.res,lwd=2)
qqline(dev.res)
```

Residual versus fitted value plot code for the PCAR model.

```
res<-c(0.9285,0.1575,-1.707,-2.049,0.3108,-1.224,1.819,0.6269,1.128)
fitted<-c(2.846,1.814,3.819,3.459,2.572,2.188,2.134,1.195,2.704)
```

```
plot(fitted,res,main="Residuals versus Fitted values",ylab="Residual",
xlab="Fitted Value",lwd=4,type='p',cex=.5)
```