MULTISCALE SPATIAL MODELLING OF DIABETES AND HYPERTENSION IN NAMIBIA

A THESIS

SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN APPLIED STATISTICS & DEMOGRAPHY

OF

THE UNIVERSITY OF NAMIBIA

BY

TOMMY RENNIEK HARRIS

200303350

APRIL 2017

Supervisor: Professor Lawrence Kazembe

ABSTRACT

In Namibia, non-communicable diseases are on the increase. Statistics on non-communicable diseases (cancer, diabetes, cardiovascular diseases, hypertension etc.) as a cause of morbidity and mortality indicate that it is a public health concern although population based estimates in the area are lacking. The Ministry of Health and Social Services stated that between 2004 and 2008, hospital based mortality due to cancer (all types) increased from 3.2% to 54.7%, cardiovascular diseases (all types) increased from 5.3% to 21.2% while diabetes mellitus also increased from 1.0% to 14.6%.

To curb the rising trend of the burden of non-communicable diseases in Namibia, the Ministry of Health and Social Services embarked on several preventive initiatives such as raising awareness through preventive programmes, passing laws related to the use of tobacco products, developing national promotion policies as well as health promoting school initiatives. The programmes however are implemented at national level but efficient targeting of such programmes requires the identification of high risk areas of diseases to identify where the disease is most prevalent. The use of disease maps to identify areas of elevated risk for non-communicable diseases in Namibia therefore becomes important with the available limited resources. Disease mapping is one such approach that can serve as a basic tool in planning to optimize the reduction of non-communicable diseases. Furthermore, mapping of such diseases allows the study of one disease at a time or multiple disease for comprehensive programming.

The study follows a quantitative cross sectional study design using multiscale disease modelling methods to describe areas of elevated risk at region, health district and constituency level in Namibia. The main aim of the study was to fit a multiscale model

to identify spatial variations of diabetes and hypertension at various geographic levels in Namibia to guide better planning, monitoring and evaluation and assist in targeting of resources. The specific objectives were to; Estimate disease risk at health district, region and constituency level in Namibia; Estimate macro determinants of diabetes and hypertension; and explore various approaches to model fitting of diabetes and hypertension in Namibia.

The population for the study was all persons using health facilities from 2008-2014 in Namibia. A total of 15462 cases of diabetes and 30620 cases of hypertension were reported from 444 health facilities over a period of seven years. The covariates considered for the study were safe water, wood/charcoal, main source of income: wages and salaries, main source of income: pension and education attainment: incomplete primary education. Covariates for the study were obtained from the 2011 Namibia Population and Housing Census.

The random effects were modelled using a conditional autoregressive prior distribution. A Fully Bayesian Inference based on Markov Chain Monte Carlo simulation techniques was used to overcome difficulties in calculating the posterior distribution.

Results from the study showed that significant spatial variation exist for both diabetes and hypertension among the different levels considered in the study. The relative risk for diabetes was found to be highest in Kavango region suggesting an increased risk for diabetes in the area. At health district and constituency level, the relative risk for diabetes was found to be highest in Rundu health district and Rundu Urban

constituency respectively suggesting that the was an increased risk of diabetes in the area. With respect to hypertension, the relative risk of hypertension was highest in Khomas region. At health district level, Andara, Rundu, Swakopmund and Windhoek had the highest relative risk indicating that there was an increased risk of hypertension in the areas. At constituency level, Rundu Urban constituency was found to exhibit the highest relative risk of hypertension suggesting an increased risk for hypertension in the areas.

For the variance components, region had the highest posterior mean (1.0168 and 1.0262 for diabetes and hypertension respectively) suggesting that diabetes and hypertension was high among the different regions in Namibia compared to health districts and constituencies.

It is hoped that the study will assist policy makers especially those involved in health planning to develop comprehensive programmes or targeted interventions in areas that were found to have elevated risk of diabetes and hypertension, in turn developing programmes and strategies aiming at improving the health and well-being of the population. Moreover, the study hopes that considering spatial factors in planning of health programmes related to diabetes and hypertension could assist in the achievement of National Development Goals such as those outlined in NDP 4 in line with the progress towards achieving international goals such as UN Millennium Development Goals.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like first and foremost to thank The Lord for allowing me the time and giving me the ability to make this a reality. Without his help throughout the research, none of this would be accomplished. To my research supervisor Professor. Lawrence Kazembe thank you very much for your guidance, advice, and inspiration throughout the entire research process. I wish also to express my sincere gratitude to my family**,** friends and fellow Masters classmates for cheering me on. Many more thanks to the Ministry of Health and Social Services for allowing me carry out this research study using their HIS data.

DEDICATION

I dedicate this work to my late grandmother who was a pillar in my life. Also dedicate this work to my friends and family who have been nothing but helpful through the entire research.

DECLARATIONS

I, Tommy Harris, hereby declare that this study is a true reflection of my own research, and that this work, or part thereof has not been submitted for a degree in any other institution of higher education.

Tommy Harris Date

…… …….

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
ALL	Acute Lymphoblastic Leukaemia
BIC	Bayesian Information Criterion
CAR	Conditional Autoregressive
CVDs	Cardiovascular diseases
DIC	Deviance Information Criterion
EB	Empirical Bayes
FB	Fully Bayesian
GIS	Geographic Information System
GLMM	Generalized Linear Mixed Models
GMCAR	Generalized Multivariate Conditional Autoregressive
GMRF	Gaussian Markov Random Field
HIS	Health Information system
IAR	Intrinsic Autoregressive
INLA	Integrated Nested Laplace Approximation
LMICs	Lower and Middle Income countries
MCAR	Multivariate Conditional Autoregressive
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
MoHSS	Ministry of Health and Social Services
MRF	Markov Random Field

NCDs	Non Communicable Diseases
NDPs	National Development Plans
NDHS	Namibia Demographic Health Survey
NSA	Namibia Statistics Agency
PH	Public Health
PQL	Penalized Quasi-Likelihood
RR	Relative Risk
SIR	Standardized Incidence Ratio
SMR	Standardized Morbidity Ratio
SSA	Sub-Sahara Africa
T1D	Type 1 diabetes
UK	United Kingdom
WHO	World Health Organisation

CHAPTER 1: INTRODUCTION

1.1 Background

Non-communicable diseases (NCDs) are increasingly becoming a challenge to achieving global progress in improving population health (Ministry of Health and Social Services [MoHSS]) & ICF International, 2014). The four main types of NCDs are cardiovascular diseases (like heart attacks and stroke), cancers, chronic respiratory diseases (including chronic obstructed pulmonary disease and asthma) and diabetes (Micha et al., 2014). The World Health Organisation [WHO] (2014) states that NCDs affect low and middle income countries, where nearly three quarters of NCD deaths – 28 million – occur. They make the largest contribution to mortality globally and account for 60 percent (35 million) of global deaths (MoHSS & ICF International, 2014). Globally, the NCD burden will increase by 17 percent in the next ten years, and in the African region by 27 percent (Islam et al., 2014).

In Sub Sahara Africa (SSA), NCDs are already a significant problem, causing almost a third of deaths in the region as a whole (Marquez &Farrington, 2013). According to Marquez and Farrington (2013), NCDs are already the leading cause of death in SSA. Moreover, the largest relative increase in NCD deaths globally in the next decade is expected to occur in Africa (Marquez & Farrington, 2013).

In Namibia, NCDs are on the increase. Statistics on NCDs (cancer, diabetes, cardiovascular diseases, hypertension etc.) as a cause of morbidity and mortality show that it is a growing public health concern even though population-based data estimates in the area are lacking (WHO, 2010; MoHSS & ICF International, 2014). The WHO (2010) stated that between 2004 and 2008, institutional mortality due to cancer (all

types) rose from 3.2% to 54.7%, cardiovascular diseases (all types) rose from 5.3% to 21.2%, and diabetes mellitus rose from 1.0% to 14.6%. Moreover, 3650 new cases of diabetes were recorded in 2010 and 2011(WHO, 2011). The most prevalent NCDs in Namibia are cardiovascular diseases which accounted for 38 percent of total deaths across all age groups in 2008 while cancers non-communicable variants of respiratory diseases, and diabetes contributed 3% to total mortality (MoHSS & Macro International, 2008).

With the rising trend in the burden of NCDs in Namibia, more efforts need to be done to mitigate the burden of NCDs. Disease mapping is one such approach that can serve as a basic tool in planning to optimize the reduction of NCDs (Lawson, 2008).

The study of spatial variations in disease rates (disease mapping) is a classic epidemiological technique, where location is used as a surrogate for the mix of lifestyle, environmental and possibly genetic factors that may underline geographical differences in risk (Held & Best, 2001). According to Held and Best (2001) and Held Natário, Fenton, Rue and Becker (2005) the purpose is to describe such variations so that areas of elevated risk can be identified. Bayesian hierarchical models are typically used in this context, which represent the risk surface using a combination of available covariate data and a set of spatial random effects. The random effects are typically modelled by a conditional autoregressive (CAR) prior distribution. In studying NCDs in Namibia, a multiscale modelling approach is proposed. The multiscale approach involves spatial analysis at different scales of aggregation, the approach is used when it is of interest to consider a relationship at different aggregation levels (Lawson,

2008). The need to study disease at different levels is important because it allows inferences to be made both at the higher (areas) and lower levels (subareas) simultaneously using a Bayesian convolution model (Aregay, Lawson, Faes & Kirby, 2014). Moreover, because a public health district contains at least one health facility or may belong to a constituency, there may be a grouping effect, i.e. health facilities belonging to the same health district or health district belonging to the same constituency may behave similarly (Aregay et al., 2014).

This study therefore seeks to apply statistical methods for disease mapping to investigate geographical variations in NCD risk at various spatial levels.

1.2 Statement of the problem

To curb the rising trend of burden of NCDs in Namibia, the MoHSS has embarked on a number of preventive initiatives including raising awareness through preventive programmes, passing laws related to the use of tobacco products as well as national promotion policy and health promoting school initiatives. These programmes are implemented at national level but efficient targeting of such programmes requires the identification of high risk areas where the disease is most prevalent. The use of disease maps to identify areas of elevated risk of NCDs burden therefore becomes important with available limited resources in Namibia.

There is however less information available on mapping of NCDs in Namibia, a disadvantage in the spatial epidemiology framework which assist to know at regional and local level the disease risk to better deploy limited resources at targeted areas in

which the risk is high. Multiscale modelling provides an opportunity to study such diseases by considering associated factors using location.

1.3 The aim of the study

The aim of the study is to fit a multiscale model to identify spatial variations of diabetes and hypertension at various geographic levels in Namibia to guide better planning, monitoring and evaluation and assist in targeting of resources.

The specific objectives:

1. To estimate diabetes and hypertension risk at region, health district and constituency level in Namibia.
2. To estimate macro determinants of diabetes and hypertension in Namibia.
3. To explore various approaches to model fitting of diabetes and hypertension in Namibia.

1.4 Significance of the study

Recent developments of epidemiological methods to study diseases risk remain of vital importance. Knowledge of disease burden in populations is essential for health authorities which seek to use limited resources to the best effect by identifying priority health programmes for prevention and care. In the same vein, the increasing power in analyzing geo-referenced data provided by Geographic Information Systems (GIS) results in a large demand for appropriate statistical data analysis. The GIS diffusion reflects in a growing utilization of disease mapping techniques for decision making

processes. The study is therefore significant as it will assist the MoHSS to derive health metrics, guide intervention strategies and advance epidemiological understanding using spatial modeling and mapping of diseases, approaches that Namibia seems not to be aware of even though strategies to combat NCDs have been developed. Moreover, there is a lack of adequate information with regard to the mapping of diseases where as other countries such as South Africa, the mapping of diseases is increasingly being undertaken to guide intervention strategies, and advance epidemiological understanding.

The growing concern about the prevalence and incidences of NDCs in Namibia motivates the study, largely because there is need to identify areas where people are at risk of diabetes and hypertension to better guide planning, monitoring and evaluation and resource allocation with considerable applications of disease mapping in studying health problems in Sub Saharan African countries having appeared over years.

1.5 Organisation of thesis

The rest of the study is structured as follows. Chapter 2 provides a review of literature on disease mapping models relevant in studying NCDs. Chapter 3 deals with topics such as how data was collected, descriptive statistics for the study and the multiscale models fitted for diabetes and hypertension. The results are reported in chapter 4 and finally discussion of the results and drawing of conclusions are presented in chapter 5.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Disease mapping is a method used by epidemiologists, medical demographers and biostatisticians to understand the geographical distribution of a disease (Meza, 2003). Disease maps may be useful for government agencies to allocate resources or identify hazards related to disease (Meza, 2003; Aregay et al., 2014). Moreover, faster computers have led to advances in disease mapping, allowing for more complex models and estimation methods, larger data sets and improved graphics (Meza, 2003).

According to Lawson, Browne and Rodeiro (2003) and Aregay et al. (2014) "the main goal is to study the distribution of disease spatially". This is supported by Lawson et al. (2000) who state that "The main aims of disease mapping are to describe geographic variation of diseases risk, suggest possible risk factors that may explain variation (hypothesis generation), identify unusual high risk areas so that appropriate actions may be taken, assess health inequalities in order to better allocation of health care resources and finally construction disease atlas". Therefore, the interest is to identify areas that have a higher risk for a certain infection (Aregay et al., 2014).

Many statistical approaches to spatial analysis fall into two categories: cluster detection or disease mapping (Gangnon & Clayton, 2003). The former adopts the hypothesis testing framework, testing the null hypothesis of a common disease rate across the study region against a clustering and the latter uses Bayes or empirical Bayes methods to produce smoothed estimates of the specific disease rates suitable for

mapping (Gangnon & Clayton, 2003). Mapping methods produce stable estimates by borrowing strength from neighbouring sub regions (Rao, 2003; Torabi, 2014). These are most useful for capturing gradual, regional changes in disease rates, and are less useful in detecting abrupt, localized changes indicative of hot spot clustering (Gangnon & Clayton, 2003).

The models proposed by Besag, York and Mollie (1991) and Waller, Carlin, Xia and Gelfand (1997) incorporate both spatially structured (spatial correlation) and unstructured (extra-Poisson variation) heterogeneity in a single model.

A few Bayesian approaches directly address the disease clustering problem, including Lawson and Clark (1999), Gangnon and Clayton (2000), Knorr-Held and Raer (2001) and Denison and Holmes (2001).

According to Gangnon and Clayton (2003), Lawson (1995) proposed a point process model for detection of cluster locations when exact case (and control) locations are known. Moreover, Lawson (2000) described an extension of this model to incorporate both localized clustering and general spatial heterogeneity of disease rates. Lawson and Clark (1999) describe the application of a point process clustering model to case count data through data augmentation. To apply their model, one imputes locations for each member of the population at risk, typically by assuming a uniform spatial distribution within each cell, to produce a point process. One then proposes a clustering model for the point process.

Gangnon and Clayton (2000), Knorr-Held and Raer (2001) and Denison and Holmes (2001) each considered a relatively nonparametric Bayesian framework for cluster detection in which cells are grouped into clusters. A single, common rate for cells belonging to the same cluster is assumed. The prior specification of Gangnon and Clayton (2000) assumes a large background area and a small number of clusters; the prior probability of a specific set of clusters is based on geographic characteristics of the clusters such as their size and shape. The prior specifications of Knorr-Held and Raer (2001) and Denison and Holmes (2001) assume clusters are defined by a set of cells chosen as cluster centers (cells belong to the cluster associated with the nearest cluster center); the prior probability of a set of clusters is based on a uniform selection of each cell as a cluster center.

All three methods provide very flexible specifications of clusters. The approaches of Knorr-Held and Raer (2001) and Denison and Holmes (2001) have some analytic advantages, while the approach of Gangnon and Clayton (2000) more directly models the prior probability of particular clusters. None of these methods include a spatial heterogeneity component in their model.

The rest of this chapter focuses on reviewing different models suitable in disease mapping as literature available suggests.

2.2 Standardized Morbidity Ratio Models (SMR)

According to Aregay et al. (2014) several authors have studied the risk using a standardized morbidity ratio (SMR). As a first step to assess the status of geographic

dispersion of disease rates, it is convenient to compute and map the SMR or Standardized Incidence Ratio (SIR) defined as the ratio of observed counts to expected counts in each region (Tzala & Best, 2007). These unbiased estimators of relative risk (RR) are commonly used in disease map presentation but suffer from many drawbacks (Lawson et al., 2000).

To estimate the RR, suppose that the country (or the region) used for disease mapping is divided into m non-overlapping small areas. Let θ_i be the unknown RR in the ith area. A direct (or crude) estimator of θ_i is given by the SMR

$$\theta_i = \frac{y_i}{e_i} \tag{1}$$

where y_i and e_i denote the observed and expected number of deaths (cases) over a given period ($i = 1,2, \dots m$) respectively. The expected number of deaths is defined as

$$e_i = n_i \left(\frac{\sum_i y_i}{\sum_i n_i}\right) \tag{2}$$

where n_i is the number of persons at risk in the ith area, and then treated as fixed.

A common assumption in disease mapping is that $y_i|\theta_i \sim_{iid} Poisson(e_i\theta_i)$. Under this assumption, the maximum likelihood (ML) estimator of θ_i is the SMR, $\theta_i = \frac{y_i}{e_i}$. However, a map of crude rates $\{\theta_i\}$ can badly distort the geographical distribution of disease incidence or mortality because it tends to be dominated by areas of low population, e_i exhibiting extreme SMR's that are least reliable.

$$Var(\theta) = \frac{y_i}{e_i} \tag{3}$$

Meza (2003) stated that since it is based on a ratio estimator, the mean and variance of SMR are highly dependent upon e_i. In addition, SMR could be very large in areas where the expected numbers of cases are small and small for areas where the expected numbers of cases are large. Furthermore, in areas where there are no observed count data or cases, the SMR is zero. This makes the interpretation of SMR difficult and it should be done with caution. Moreover, (Aregay et al., 2014) adds that the approach is simple and does not accommodate the correlation between neighbours.

2.3 The Besag,York & Mollie (BYM) Model

Besag et al. (1991) proposed a convolution model that allowed the RR to be statistically modelled by including spatially structured and unstructured random effects into the model (Aregay et al., 2014). This was supported by Tu and Greenwood (2012) who stated that bayesian methods were proposed to deal with sparse data arising, for instance, through small incidence or mortality rates within the context of an ecological analysis. Moreover, Tu and Greenwood (2012) emphasised that this approach improves the precision and stability of risk estimates. In addition, the method also provide a framework to model simultaneously the spatial and non-spatial (or heterogeneity) effects on disease risk.

The principles underlying using Besag's statistical model is that it allows to differentiate between the relative contribution of the spatial and non-spatial effects on disease risk. Moreover, the non-spatial or heterogeneity random effects appear in the model as extra-Poisson variation. The non-spatial effects usually arise through the variation among the populations at risk due to omitted covariates. The model assumes

that the spatial random effects control for unmeasured spatial covariates and the spatial effects are assumed to be similar across close or adjacent geographical areas.

According to Tu & Greenwood (2012), the model is defined such that the observed disease counts O_i in each area i, with associated expected counts E_i, are assumed to take a Poisson distribution i.e. $O_i \sim Poisson(E_i RR_i)$ *for* $i = 1, \dots, N$ areas, where RR_i is the relative risk of disease in area i. The maximum likelihood estimate of the RR of disease in area i equals $\frac{O_i}{E_i}$, which is the SIR_i.

Referring to the model of Besag et al. (1991), the logarithm of the RR for each area i is modelled so that:

$$log(RR_i) = \alpha_0 + \beta^T x_i + u_i + v_i \quad (4)$$

where α_0 represents the intercept of the log RR; x_i represents the covariate for each area with associated parameter β; u_i represent the independent heterogeneity effects between areas and are synonymous with extra-Poisson variation; v_i represent the spatially dependent random effects which are defined by a range of different structures describing adjacency or closeness in space (Clayton & Kaldor, 1987).

This class of models is referred to as convolution models and generally we may define a normal prior distribution for the non-spatial heterogeneity effects such that $u_i \sim Normal(0, \sigma_u^2)$ (Besag et al., 1991). We may assume that the spatially correlated random effects v_i arise through a combination of independent random effects errors e_i that are normally distributed i.e. $e_i \sim Normal(0, \sigma_e^2)$ (Tu & Greenwood, 2012). Here we assume that the components v_i may be written as

$$v_i = \frac{\sum_{j \in \Theta_i} e_j}{n_i} \quad (5)$$

Where Θ_i represents the set of areas sharing a common boundary with area i, with n_i denoting the number of neighbours for area i. Therefore, through the averaging of the independent random effects, e_j defines the effect the area j on the disease risk in area i.

The Besag model can be extended to include area-level covariates such as socio-economic status, whilst accounting for both the spatially structured and unstructured (heterogeneity) random effects for the RRs across areas.

The Besag model has its drawbacks, for instance, the model does not accommodate the spatial scaling effects associated to the aggregations of the data space. The scaling effect in disease mapping is of special interest when dealing with multiple diseases and multiscale various level diseases (Aregay et al., 2014). Moreover, this is supported by Louie and Kolaczyk (2006) who asserts that the concept of scale determines greater or lesser degree of aggregation of the information underlying the data and the effects of scale on the analysis and inference of geo-spatial data is often referred to collectively under the modifiable areal unit problem. Kolaczyk and Haung (2001) developed a multiscale modelling approach by factorizing the likelihood into individual components of local information to encompass the scaling effect.

2.3.1 Models for spatial autocorrelation in disease mapping

CARs models are commonly used to represent spatial autocorrelation in data relating to a set of non-overlapping areal units, which arise in a wide variety of applications including agriculture, education, epidemiology and image analysis. Such models are

typically specified in a hierarchical Bayesian framework, with inference based on Markov chain Monte Carlo (MCMC) simulation (Lee & Kim, 2008).

The general goal of these spatial models is to unveil and quantify spatial relations present among the data, in particular, to quantify how quantities of interest vary with explanatory variables and to detect clusters of hot spots (Oliveira, 2012).

The CAR models are typically specified as a prior distribution for a set of random effects, as part of a hierarchical Bayesian model (Lee & Mitchel, 2013). In the public health set-up, the CAR models are used to study regional patterns of the disease (Torabi & Rosychuk, 2012). The CAR models are most appropriate in the univariate case when mapping a single disease is our interest. Moreover, if information about multiple diseases as multivariate areal data is available, one may use separate univariate CAR model for each disease (Torabi, 2014).

However, if a number of diseases share the same set of spatial risk factors or the diseases may be related to one another, it is may be appropriate to use multivariate areal models to properly analyse the data (Jin, Carlin & Banerjee, 2005). Multivariate models enable us to model dependence among the multivariate components while maintaining spatial dependence between regions (Jin et al., 2005; Torabi, 2014). Section 2.3.1.1 and section 2.3.1.2 discusses univariate and multivariate models used in disease mapping.

2.3.1.1 Univariate CAR modeling

Jin et al. (2005) described the CAR model for the univariate case by first considering a univariate spatial random variable $\emptyset_i$ observed at n area locations, and define $(\emptyset = \emptyset, \dots, \emptyset_n)'$. Under the multivariate normal Markov random field (MRF)

assumption, a n full conditional distributions can then be specified as $p\left(\emptyset_i \middle| \emptyset_j, j \neq i, \tau_i^{-1}\right)$

$$= N\left(\alpha \sum_{i\sim j} b_{ij\phi_j}, \tau_i^{-1}\right), \quad i, j = 1, \ldots, n, \tag{6}$$

where $i\sim j$ denotes that region j is a neighbor (defined in terms of spatial adjacency) of region i. In addition, Jin et al. (2005) state that according to the Hammersley-Clifford theorem and brook's Lemma, the full conditional distributions in (6) uniquely determine the joint distribution,

$$\phi\sim N(0, \left[D_{\tau(I-\alpha B)}\right]^{-1}), \tag{7}$$

where B is an $n \times n$ matrix with $b_{ii} = 0$, *and* $D_\tau = Diag(\tau_i)$; usually we assume that $D_\tau = \tau D$, where D is an $n \times n$ diagonal matrix. α is a smoothing parameter, and is often interpreted as measuring spatial association.

From the CAR formulation α, D, and B can be choosen such that various CAR model structures can be obtained. The most popular CAR implementation is the pairwise difference formulation also known as the Intrinsic Auto Regressive (IAR) model (Besag et al., 1991).

When using IAR, the smoothing parameter is set to 1, i.*e* α=1 and also typically take $D = Diag(m_i)$, where m_i is the number of neighborus of region i, and $B = D^{-1}W$. Where W denote the adjacency matrix of the map, B is called the scaled adjacency matrix in this case. The CAR formulation then becomes:

$$\phi\sim N(0, [\tau(D - W)]^{-1}). \tag{8}$$

2.3.1.2 Multivariate CAR modelling

Most multivariate CAR models are members of the family developed by Mardia (1998). There are several multivariate areal models in the literature. Mardia (1988) considered the theoretical background for multivariate normal MRF specification (Jin et al., 2005; Torabi, 2014).

Billheimer et al. (1997) studied a hierarchical statistical model for compositional monitoring data using a multivariate MRF in a state-space setting. A twofold CAR model for counts of two different diseases over each areal unit was developed by Kim, Sun and Tsutakawa (2001).

A multi objective version of the CAR model which allows for flexible modelling of the spatial dependence structure was proposed by Sain and Cressie (2002). Multivariate conditional autoregressive (MCAR) models for hierarchical modelling based on MRF have been also developed (Carlin & Banerjee, 2003).

Jin et al. (2005) also introduced a flexible class of generalized multivariate conditional autoregressive (GMCAR) models for areal data using the Bayesian approach. The method of Jin et al. (2005) directly specifies the joint distribution for a multivariate Markov Random Fields (MRF) through the specification of simpler conditional and marginal models.

Analogous to the univariate case, the joint distributions are derived from the full conditional distributions (Jin et al., 2005; Torabi, 2014). Under the MRF assumption, the conditional distributions can be specified as:

$$p\left(v_i \middle| v_j \neq i, \Gamma_i^{-1}\right) = N\left(R_i \sum_{i\sim j} B_{ij} v_j, \Gamma_i^{-1}\right), i, j = 1, \dots, n, \quad (9)$$

where $v_i = \left(\emptyset_{i1}, \emptyset_{i2}, \dots, \emptyset_{i_p}\right)'$ is a p-dimensinal vector, and $\Gamma_i, R_i,$ and B_{ij} are p x p matrices. Mardia (1988) proved that the full conditional distributions in (9) uniquely determine the joint distribution

$$v \sim N(0, [\Gamma(I - B_R)]^{-1}), \quad (10)$$

where $v' = (v_1', v_2', \dots, v_n')$, B_R is a np x np matrix with $(B_R)_{ij} = R_i B_{ij}, (B_R)_{ii} = 0,$ and Γ is an np x np block diagonal matrix with p x p diagnonal entries Γ_i. From the CAR formulations (10), different Γ and B_R matrices can be chosen to obtain different MCAR model structures (Torabi, 2014).

To obtain proper joint distribution, $\Gamma(I - B_R)$ should be a positive definite and symmetric matrix, but establishing these conditions can be difficult in some cases. The formulation can be simplified by assuming that $R_i = \alpha I_{pxp}$ for $i = 1, \dots, n$ (where α is a smoothing parameter), and $\Gamma = D \otimes \Lambda$.

The model thus becomes,

$$v \sim N(0, [(D(I - \alpha B)) \otimes \Lambda]\hat{} - 1) \quad (11)$$

where Λ is a p x p positive definite and symmetric matrix.

Different MCAR models can be obtained, for example the IAR model to the multivariate case can be obtained by setting α=1. The MCAR formulation thus becomes

$$v \sim N(0, [(D - \alpha W) \otimes \Lambda]^{-1}) \quad (12)$$

assuming that $D = Diag(m_i)$, the scaled adjacency matrix $B = D^{-1}W$, and taking $\alpha \epsilon (-1,1)$. This model is denoted as MCAR (α, Λ) in Carlin and Banerjee (2003).

All of the above MCAR models are generalized from univariate CAR models under the assumption that $R_i = \alpha I_{pxp}, i = 1, \dots, n$, and can be used for any dimension *p*.

2.4 Multiscale models

The scaling effect is a special interest in disease mapping (Aregay et al., 2014). Kolaczyk and Haung (2001) developed a multiscale modelling approach by factorizing the likelihood into individual components of local information to encompass the scaling effect. A drawback when using the Besag model since the model does not accommodate the spatial scaling effects associated to the aggregations of the data space (Besag et al., 1991).

According to Kolaczyk and Haung (2001), the Besag model (1991) assumes that the hierarchical partitions correspond to successive aggregation of an initial data space. Moreover, the approach assumes the effects at the higher level are fixed not random, therefore the Besag model is not flexible enough to make inference both at the higher and lower level at the same time. The analyses of data at different resolution levels is implemented when the interest is to include all data at levels of aggregation in an analysis of all the levels (Lawson, 2008). Louie and Kolaczyk (2006) gave an example of multiple scale analysis for disease data.

For multiscale analysis, assume that health facility data and health district data of a disease is observed. The health facility is a unique subdivision of the district set i.e., each health facility falls uniquely within one health district. Since public health districts are administrative units within which certain health services are provided, it is possible that some grouping effect based on health district could be found for health facilities that lie within a given district. Hence there is multiscale information which is completely aligned in the sense that the lower level health facilities fall completely and uniquely within the higher aggregation levels units (health districts in Namibia).

To accommodate the scaling effect during modeling, Louie and Kolaczyk (2006) proposed to factorize the likelihood which contains the information of the scaling effect in a multiscale fashion under the assumed Poisson model. , Louie and Kolaczyk (2006) assumed a multinomial distribution for the data at the finer level (for instance, data at health facility level) conditioning on the coarser level (for instance, data aggregated to the health district level or region level).

Suppose the data set is aggregated at the county and public health district levels. Let $y_i^c, i = 1, \dots, N$, be the county level count of disease and $y_j^{ph} = \sum_{i \epsilon j} y_i^c, j = 1, \dots, n$, *is* the j^{th} public health district level. *Let* $y_i^c \sim Poisson(e_i^c \theta_i^c)$ and $y_j^{ph} \sim Poisson\left(e_j^{ph} \theta_j^{ph}\right)$ then the model is given by

$$log(\theta_i^c) = \alpha_0^c + v_i^c + u_j^{ph}, \quad (13)$$

$log\left(\theta_j^{ph}\right) = \alpha_0^{ph} + v_j^{ph} + u_j^{ph}$ respectively. (14)

Here, e_i^c and e_j^{ph} are the expected rate at the county and public health district level respectively. $\alpha_0^c \; and \; \alpha_0^{ph}$ are intercept parameters, v_i^c and v_j^{ph} are the uncorrelated heterogeneity (UH) random effects and u_j^{ph} is a shared spatial structured random effect.

Aregay et al. (2014) used Bayesian multiscale modelling for aggregated data to study oral cancer in Georgia. Aregay et al. (2014) proposed 3 different models that encompassed the scaling effect. In addition the models were assessed using the deviance information criterion (DIC) to find out how the different models fit the data.

2.5 Model parameter estimation

The estimation of RR involves generalized linear mixed models (GLMM) which implies the prediction of random effects that represents the RRs (Ugarte, Militino & Goicoa, 2008).

A few approaches for parameter estimation of the models in disease mapping are discussed sections 2.5.1-2.5.4. Empirical Bayes (EB) and a Fully Bayesian (FB) approaches are usually used for smoothing purposes (Ugarte et al., 2008). For maximum likelihood estimation of GLMM with counts, numerical integration is usually required. The penalized quasi-likelihood technique (PQL), a Laplace approximation to the quasi-likelihood can reduce the problem to a series of weighted least squares regressions (Breslow and Clayton, 1993).

2.5.1 Fully Bayesian Approach (FB)

The FB approach in disease mapping exploits the posterior distribution of the RRs to obtain reliable and smooth estimates (Ugarte et al., 2008). The approach became popular because of its flexibility and the availability of free software implementing MCMC methods allowing the fit of complex models (Ugarte et al., 2008). One of the main advantages of the FB approach is that it provides, for each area, samples of the whole posterior distribution of the RRs, supplying more information than just a single point estimate.

In a FB setting, all the parameters are assigned prior distribution whose parameters are also assigned hyperprior distributions to cope with their possible variability. The estimation procedures can be done by using MCMC simulation techniques. Best, Richardson and Thomson (2005) summarised the most recent Hierarchical Bayesian models that are used for disease mapping using Full Bayes estimation.

Since the FB inference is based on the analysis of posterior distribution of the model parameters. The posterior distribution based on conditional independent assumption is given by:

$$p(\beta, \tau^2, \gamma|data) \propto L(data|\beta, \tau^2, \Upsilon)\, p(\beta, \tau^2, \Upsilon)$$

$$= L(data|\beta, \tau^2, g)x \left\{\prod_{i=1}^{\infty} p\left(\beta_j\middle|\tau^2\right) p\left(\tau_j^2\right)\right\} p(\gamma) \qquad (15)$$

whereby, $p(\beta, \tau^2, \gamma|data)$ is the posterior density function while $L(data|\beta, \tau^2, \Upsilon)$ gives the likelihood for the data.

The spatial interaction is usually modelled by means of a CAR model (Waller & Gotway, 2004), so that only the effects of nearby areas are included.

2.5.2 Empirical Bayes approaches (EB)

Most of the statistical techniques for smoothing the RRs use EB method. The basic idea from Efron and Morris (1975) is pooling the information across the regions through a suitable model using James-Stein (1961) estimators (Gómez-Rubio & López-Quılez, 2006).

In the EB approach, the posterior distributions of the parameters of interest given the data are first obtained assuming the model parameters are known. The model parameters are then estimated by suitable methods and inferences are made from the estimated posterior distributions (Ugarte, Goicoa & Militino, 2009).

EB approaches can be found in studies conducted by Clayton and Kaldor (1987), Cressie and Chan (1989) and Cressie (1992) using Poisson likelihood, log-normal prior and the spatial modeling of log-RRs.

2.5.3 Maximum Likelihood Estimation (MLE)

When the data has been observed, the likelihood function, or simply the likelihood is constructed. The likelihood is the joint probability function of the data, but viewed as a function of the parameters, treating the observed data as fixed quantities. Assuming that the data values, $\boldsymbol{x} = (x_1, \ldots, x_n)$ are obtained independently, the likelihood function is given by

$$L(\theta;\, y) = p(y_1, \ldots, y\_n|\theta) = \prod_{i=1}^{n} p(y_i|\theta) \qquad (16)$$

In the Bayesian framework, all information about θ directly from the data is contained in the likelihood. Values of the parameters that correspond with the largest values of the likelihood are the parameters that are most supported by the data. To obtain the posterior distribution, $p(y_i|\theta)$ the probability distribution of the parameters once the data have been observed, the Bayes' theorem is applied such that:

$$p(\boldsymbol{\theta}|y) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})}{\int p(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})d\theta} = \frac{p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{y})}{p(y)} \propto p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{y}) \qquad (17)$$

where ∝ means "is proportional to" (i.e., that the expressions are equal when the right-most term is multiplied by a normalizing constant that doesn't depend on $\boldsymbol{\theta}$). Operationally, therefore, it is straightforward in principle to obtain the posterior distribution: Simply multiply the prior distribution by the likelihood and then determine the constant (not depending on $\boldsymbol{\theta}$) that forces the expression to integrate to 1. An effective strategy for computing the posterior distribution is to drop multiplicative constants from the prior distribution and likelihood that do not depend on *q*, and then in the final step determine the normalizing constant.

2.5.4 Integrated Nested Laplace approximation (INLA)

Literature suggests that the Integrated nested Laplace approximation (INLA) is mostly used in implementing Approximate Bayesian Inference (Martino & Rue, 2010). Moreover, Martino and Rue (2010) state that INLA is a new approach to statistical inference for Latent Gaussian Markov random field (GMRF) models introduced by Rue, Martino and Chopin (2009). According to Martino and Rue (2010) it provides a

fast, deterministic alternatives to MCMC that is at the moment the standard tool for inference in GMRF models.

A Latent GMRF model is a hierarchical model where, at the first stage a distributional for the observables y usually assumed to be conditionally independent given some latent parameters η and, possibly, some additional parameters $\theta_1$is found.

$$\pi(y|\eta, \theta_i) = \prod_j \pi\left(y_j \middle| \eta_j, \theta_1\right) \quad (18)$$

The latent parameters η are part of a larger latent random field x, which constitutes the second stage of our hierarchical model. The latent field x is modelled as a GMRF with precision matrix Q depending on some hyperparameters θ_2.

$$\pi(x|\theta_2) \propto exp\{-\frac{1}{2}(x-u)^T Q(x-u)\} \quad (19)$$

The third, and last, stage of the model consists of the prior distribution for the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$**.**

The INLA approach provides a recipe for fast Bayesian Inference using accurate approximations *to* $\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ and $\pi(xi|y), i = 0, \ldots, n-1$, i.e. the marginal posterior density for the hyperparameters and the posterior marginal densities for the latent variables.

Different types of approximations are available and the approximate posterior marginals can be used to compute summary statistics of interest, such as posterior means, variances or quantiles.

2.6 Model Selection Criterion

2.6.1 Introduction

The need to select a model is of great importance in statistics in order to ensure goodness of fit and adjust or penalize for model complexity. The observed data is usually from an unknown probability distribution. As a result, several models are fitted in order to find the best.

2.6.2 Akaike Information Criterion (AIC)

The AIC is a statistics used to select the best model. It is defined as

$$AIC = 2k - 2\,ln(L) \tag{20}$$

where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model.

The AIC is calculated for each model under consideration using the same data and the model with the lowest AIC is chosen. The term $2k$ is a penalty to be paid for over fitting and this discourages adding too many variables in the models which always leads to a smaller likelihood.

2.6.3 Bayesian Information Criterion (BIC)

The BIC is another model introduced by Schwarz (1978), the model selection criterion is based on the empirical log-likelihood. The model does not require the specifications of priors therefore favoured in situations where the priors are difficult to set. It is similar to the AIC and both statistics penalize model complexity.

The best model fitted is identified by the minimum value of BIC. The BIC is given by

$$BIC = -2\,ln(L) + kln(n) \quad (21)$$

where,

l= the maximum value of the likelihood function of the model and

k= the number of free parameters to be estimated

n= the number of observations, or equivalently or the sample size

The penalty term in the BIC is more stringent than the penalty team of AIC for $n > 8, kln(n)$ exceeds $2k$ and this lead to BIC favouring smaller models than the AIC.

2.6.4 Deviance Information Criterion (DIC)

The DIC (Zhu & Carlin, 2000) is a generalization of the AIC and the BIC and is widely used in model selection where MCMC simulation is used.

The deviance is given as:

$$D(\theta) = -2\,log\big(p(y|\theta)\big) + 2\log\big(f(y)\big) \quad (22)$$

whereby y are the data, θ are the unknown parameters of the model and $p(y|\theta)$ is likelihood function and $f(y) = p(y|u(\theta) = y)$. The effective number of parameters of the model is computed as

$$pD = \overline{D} - D(\bar{\theta}) \quad (23)$$

whereby $\bar{\theta}$ is the expectation of θ and pD is the effective number of parameters. Thus DIC is then calculated as

$$DIC = pD - \overline{D} \quad (24)$$

In model selection, the general rule is that models with smaller DIC be preferred over models with a larger DIC.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

The study considers using some of the methods reviewed in Chapter 2 to carry out spatial analyses of diabetes and hypertension in Namibia. The spatial analysis was done at region, health district and constituency level. The study considered disease mapping at constituency level because resources for planning in Namibia are also decentralised to the constituency level.

3.2 Study area characteristics

Namibia is a country in Southern Africa whose western border is Atlantic Ocean. It shares borders with Angola and Zambia to the North, Botswana to the East and South Africa to the South and East. Namibia was divided into thirteen administrative regions during the 2011 Housing and population census, which were further sub divided into 107 constituencies. As of 2013, Namibia has been divided into 14 Regions and 121 constituencies. See appendix B and C (Part A and Part B) for details of regions and constituencies in Namibia.

According to the Namibia Population and Housing Census conducted in 2011, the total population for Namibia was estimated to have increased to 2.1 million from 1.8 million in 2001 estimated in 2001 (Namibia Statistics Agency [NSA], 2015). A relatively large proportion (58 %) of Namibia population currently lives in the rural areas.

Namibia has a total of 34 health districts offering health care services and each health district has several health facilities. Within a constituency, at least one health facility

belongs to a certain constituency. See appendix A for details of health districts in Namibia.

3.3 Research design and Data collection

This is a quantitative cross-sectional study, using secondary administrative data sourced from the Health Information System (HIS) of the MoHSS.

Data for the study combined inpatient and outpatient observed cases of diabetes and hypertension from 2008-2014 sourced from the HIS different health facilities in Namibia. The HIS is a public health planning and information system used to manage data from all 444 health facilities in Namibia.

A number of variables are being collected from patients/clients visiting public and mission health facilities, which are used to create indicators. Indicators collected by the system are on a basis of their relevance to global public health. Taken together, these indicators provide a comprehensive summary of the current status of the national health and health systems in cause-specific mortality and morbidity, selected infectious diseases, health service coverage, risk factors and health systems. The raw data from individual patient encounters is collected each day at the facility level using paper based registers and tally sheets. At the end of each month, the data is transferred onto a monthly summary report forms, which are submitted to the district level. The variables were however not made available by the MoHSS for this analysis.

3.4 Data management

Data required to conduct the study included the identification of covariates to be added to the models during the analysis stage. The explanatory variables were obtained from the 2011 Namibia Population and Housing Census. Socio-economic explanatory variables included education attainment, household conditions and main source of income. All explanatory variables were centered about the mean during the analysis to ensure that the covariate values are not too far away from zero, that is, the covariates have neither too large negative nor too large positive values.

The outcome variables for the study were diabetes and hypertension. Total observed counts of diabetes and hypertension were aggregated to the health district level using health facility level data from 2008 to 2014.

The expected values of observed cases using the population at risk was also calculated. At district level, the population at risk considered was population aged 15 years and above. At constituency level, the population at risk considered was aged 15-64 years.

Data sorting and cleaning was carried out using Microsoft excel program to remove missing data.

For the health districts, R statistical software was used to create boundaries of health districts in Namibia. The boundaries file created by R were used to compute the neibourhood information of the map of all health districts in Namibia in BayesX and visualize results of RR at health district level. The boundaries for the map of all regions and constituencies in Namibia were readily available, this was used to visualize results

of RR at region and constituency level. The boundary maps of all health districts, regions and constituency in Namibia were also used in the different models fitted.

3.5 Ethical considerations

The permission to use the 2008-2014 diabetes and hypertension data was granted by the MoHSS.

3.6 Description of key variables

The outcome variables for this study are diabetes and hypertension counts of data aggregated to the health district, region and constituency level to carry out multiscale data analysis. In addition, socio economic variables were considered as covariates for modelling. Table 1 shows the description of the outcome variables used in the study and the covariates used to fit models in BayesX. Covariates considered for the study were safe water, wood or charcoal, main source of income: wages and salaries, main source of income: pension and education attainment: incomplete primary (for a description of the variables, Table 1).

Table 1: Description of key variables

Covariates	**Description**
Outcome variables	
Diabetes	The number of observed cases of diabetes at a health district/region/constituency
Hypertension	The number of observed cases of hypertension at a health district/region/constituency

Spatial effects	
Health district	A district that has one or more clinic or hospital or mission providing health services to the public. (See Appendix A)
Region	Administrative boundaries, there were 13 regions in Namibia. See Appendix B
Constituency	Administrative boundaries, there were 107 constituency in Namibia. See Appendix C, part A and part B.
Socio-economic factors	
Safe water	Proportion of households with safe water
Wood or charcoal	Proportion of households depended on wood or coal as source for cooking
Main source of income: Wages and salaries	Proportion of households depended on wages and salaries as main source of income
Main source of income: pension	Proportion of households depended on pension as main source of income
Education attainment: Incomplete primary	Proportion of population that have not completed primary education

3.7 Data Analysis

To perform multiscale analysis, three spatial effects were considered into the models, namely region, health district and constituency to account for spatial variation in diabetes and hypertension at region, health district and constituency level.

3.7.1 Descriptive analysis

The observed cases of diabetes and hypertension were mapped at region, health district and constituency level to visualize the distribution of diabetes and hypertension. The maps of the distribution of the 2 diseases provided a basis for comparison purposes for the SMR maps, BYM model and the multiscale models fitted.

The SMR for diabetes and hypertension at the region, health district and constituency level was then calculated and mapped as part of the descriptive analysis. The SMR was used to reduce the noise in the data aiming to produce more plausible maps.

SMR is defined as the ratio of observed disease incidence relative to what is expected under standard conditions (Clayton & Kaldor, 1987). The SMR is a crude approach and SMRs can provide unstable estimates due to their ratio form. The formula used to come up with estimates of SMR at region, health district and constituency level was given in section 2.2 of chapter 2.

3.7.2 CAR model

Whilst relatively small population totals per area result in large random variability and unstable standardized incidence ratios, Bayesian disease mapping methods usually assist in smoothing the RR estimates using the CAR model (Kazembe & Mpeketula,

2009). To address the problem with large random variability and producing unstable estimates, 3 BYM models for each disease to smooth the RR estimates at region, health district and constituency level were fitted. Spatially structured variability to generate locally smoothed RR estimates of diabetes and hypertension, which is achieved by pooling information from neighbouring areas were also assumed. The CAR model was used to smooth the data. Two areas were assumed neighbours if they have a common boundary. The study used BayesX 2.1 statistical software for inference based on MCMC simulation techniques.

3.7.3 Multiscale modelling

Multiscale models are used to describe data at different geographical levels. These methods allow the study of the relationship between predictors and outcomes at multiple scales simultaneously. The methods can assist investigating whether or not the relationship at the finer level will hold true at the coarser level. In this study, multiscale models were applied to model the relationship between spatially referenced outcomes and predictors by taking into account the spatial scaling effect.

Health facility data collected at health facility level was aggregated to the health district, region and constituency level to carry out multiscale analysis. For spatial health data, it has been recognized that a geographical correlation will exist between spatial units because spatial units close together in space often have similar disease risk, whereas regions far apart are often different. Hence, a spatially structured random effect was used to handle the association between neighbours. The spatial effects introduced into the models to account for spatial variation in diabetes and hypertension

were health district, constituency and region. The socio-economic factors considered for the study were modelled as fixed effects only. The socio-economic factors were added to the multiscale models selected based on the rule of thumb suggested by Best and Richardson (2009, 2012).

3.8 Model choice and Inference

The study considered fitting a multiscale model for diabetes first, a multiscale model for hypertension was then fitted. In both cases, health district, region and constituency were considered as spatial effects.

The candidate models fitted considered the spatially structured random effects, spatially unstructured random effects and fixed effects for the outcome variables diabetes and hypertension. For each outcome variable, eleven models were fitted (i.e. for health district, constituency and region as spatial effects, eleven models were fitted for diabetes and for hypertension as an outcome variable, eleven models were fitted while considering health district, constituency and region as spatial effects). The description of the models fitted are provided in Table 2.

With regard to the levels, the idea was to fit three random effects models at the same level (health district alone as a spatial effects or region alone as a spatial effect) and then fit another three random effects models at the same level (for instance, region only or constituency only) while considering spatially structured and spatially unstructured effects. To fit multiscale models, different random effects models were considered while the levels were different at all times (for instance, health district and constituency or health district and region).

The RR was calculated and mapped after fitting the multiscale model for diabetes and hypertension separately. A multisclae model with covariates for diabetes was used for further analysis while for hypertension, a multiscale model without covariates was used in the analysis.

A large number of models (twenty-two models in total) were fitted because the study considered two outcome variables which were fitted separately to examine spatial variation based on different variables in the model. In addition, the study considered three levels (health district, region and constituency) when modeling of outcomes were conducted because many public health policies decisions in Namibia are based on associations obtained from the analysis of data available at various levels, the various levels were therefore important to be accounted during the study. Moreover, most public health programmes in Namibia are aligned to various public health policies and are usually at the levels considered in the study.

The study applied a FB approach based on Markov Priors and used the MCMC techniques for inference and model checking. MCMC techniques assist to overcome difficulties in calculation of the posterior distribution. The techniques allows for drawing random samples from the posterior, whose characteristics (such as the mean, standard deviation and quantiles) can be easily estimated by their empirical analogues.

For model choice, the DIC developed as a measure of fit and model complexity was used (Spiegelhalter, 2002; Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2012). Generally the best model fit is given by the model that minimizes DIC, however the rule of thumb by Best and Richardson (2009, 2012) suggest a significant difference in

the model fit if the ΔDIC ≥ 4. The DIC was therefore primarily used for model comparison and rule of thumb for ΔDIC was used to select best fit model for diabetes and hypertension on which the analysis is based. Models with ΔDIC less than four (4) were considered to be similar and therefore any of the models could be selected as the best fit model to describe spatial variation of diabetes and hypertension. The study also reported on the deviance and P_D of the models fitted. Boxplots of residuals were used to measure the goodness of fit of the models fitted.

The analysis was carried out using version 2.1 of the BayesX (Belitz, Brezger, Kneib & Lang, 2009) statistical software which permits Bayesian inference based on MCMC simulation techniques. The R (R Development Core Team, 2015) software was used primarily for developing boxplots of residuals of fitted models. For all the models, 52000 iterations were run with a burn of 2000 for each model.

The following candidate models were fitted in BayesX (The description of the models fitted in BayesX is given in Table 2).

$M_1 = Fixed\ effects\ only\ for\ diabetes$

$M_2 = f_{spatial(hd)}$

$M_3 = f_{spatial(const)}$

$M_4 = f_{spatial(reg)}$

$M_5 = f_{spatial(hd)} + f_{random(hd)}$

$M_6 = f_{spatial(const)} + f_{random(const)}$

$M_7 = f_{spatial(reg)} + f_{random(reg)}$

$M_8 = f_{spatial(hd)} + f_{spatial(const)}$

$M_9 = f_{spatial(hd)} + f_{spatial(reg)}$

$$M_{10} = f_{spatial(hd)} + f_{random(hd)} + f_{spatial(const)}$$

$$M_{11} = f_{spatial(const)} + f_{random(const)} + f_{spatial(hd)} + f_{spatial(reg)}$$

$$M_{12} = f_{spatial(const)} + f_{random(const)} + f_{spatial(hd)} + f_{spatial(reg)} + covaraites$$

$$M_{13} = Fixed\ effects\ only\ for\ hypertension$$

$$M_{14} = f_{spatial(hd)}$$

$$M_{15} = f_{spatial(const)}$$

$$M_{16} = f_{spatial(reg)}$$

$$M_{17} = f_{spatial(hd)} + f_{random(hd)}$$

$$M_{18} = f_{spatial(const)} + f_{random(const)}$$

$$M_{19} = f_{spatial(reg)} + f_{random(reg)}$$

$$M_{20} = f_{spatial(hd)} + f_{spatial(const)}$$

$$M_{21} = f_{spatial(hd)} + f_{spatial(reg)}$$

$$M_{22} = f_{spatial(hd)} + f_{random(hd)} + f_{spatial(const)}$$

$$M_{23} = f_{spatial(const)} + f_{random(const)} + f_{spatial(hd)} + f_{spatial(reg)}$$

$$M_{24} = f_{spatial(const)} + f_{random(const)} + f_{spatial(hd)} + f_{spatial(reg)} + covaraites$$

Table 2: Description of the models for the study.

M_1:	Fitted the fixed effects for the outcome diabetes.
M_2:	Fitted the spatially structured random effects for the outcome diabetes at health district level.
M_3:	Fitted the spatially structured random effects for the outcome diabetes at constituency level.
M_4:	Fitted the spatially structured random effects for the outcome diabetes at regional level.

M_5:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome diabetes at health district level.
M_6:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome diabetes at constituency level.
M_7:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome diabetes at regional level.
M_8:	Fitted the spatially structured random effects at health district level plus the spatially structured random effects at constituency level for the outcome diabetes.
M_9:	Fitted the spatially structured random effects at health district level plus the spatially structured random effects at region level for the outcome diabetes.
M_{10}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome diabetes at health district level plus the spatially structured random effects of diabetes at constituency level.

M_{11}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome diabetes at constituency level plus the spatially structured random effects of diabetes at health district level plus the spatially structured random effects of diabetes at regional level.
M_{12}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome diabetes at constituency level plus the spatially structured random effects of diabetes at health district level plus the spatially structured random effects of diabetes at regional level plus fixed effects.
M_{13}:	Fitted the fixed effects for the outcome hypertension.
M_{14}:	Fitted the spatially structured random effects for the outcome hypertension at health district level.
M_{15}:	Fitted the spatially structured random effects for the outcome hypertension at constituency level.
M_{16}:	Fitted the spatially structured random effects for the outcome hypertension at regional level.
M_{17}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome hypertension at health district level.

M_{18}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome hypertension at constituency level.
M_{19}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome hypertension at regional level.
M_{20}:	Fitted the spatially structured random effects at health district level plus the spatially structured random effects at constituency level for the outcome hypertension.
M_{21}:	Fitted the spatially structured random effects at health district level plus the spatially structured random effects at region level for the outcome hypertension.
M_{22}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome hypertension at health district level plus the spatially structured random effects of hypertension at constituency level.
M_{23}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome hypertension at constituency level plus the spatially structured random effects of

	hypertension at health district level plus the spatially structured random effects of hypertension at regional level.
M_{24}:	Fitted the spatially structured random effects plus the spatially unstructured random effects for the outcome hypertension at constituency level plus the spatially structured random effects of hypertension at health district level plus the spatially structured random effects of hypertension at regional level plus fixed effects.

CHAPTER 4: RESULTS

4.1 Introduction

Chapter 4 gives results of the models fitted and the model selection criterion used. In addition, the parameter estimates for the best model fit is carried out and the interpretation of the results and the subsequent presentation of the RR of diabetes and hypertension at region, health district and constituency level into the maps are undertaken.

4.2 Distribution of observed counts of diabetes and hypertension at regional level

The observed counts of diabetes and hypertension were aggregated to the regional level to map the distribution of the observed counts at various regions. The distribution of diabetes and hypertension at regional level is shown in Figure 1 (a) and Figure 1 (b) respectively.

According to Figure 1 (a), Khomas region was found to have the highest number of observed counts of diabetes while Hardap, Erongo, Karas, Otjozondjupa, Omusati, Ohangwena and Kavango region had the lowest number of observed counts. Caprivi region also appeared to have a high number of observed counts of diabetes.

For observed counts of hypertension, Figure 1 (b) indicate that the highest number of observed counts were in Khomas region. Caprivi, Oshikoto and Omaheke also showed a darker grey color on the map suggesting a high number of observed counts in those areas as shown in Figure 1 (b). The lowest counts of hypertension was found in the southern part of Namibia, namely in Hardap and Karas region.

(a) (b)

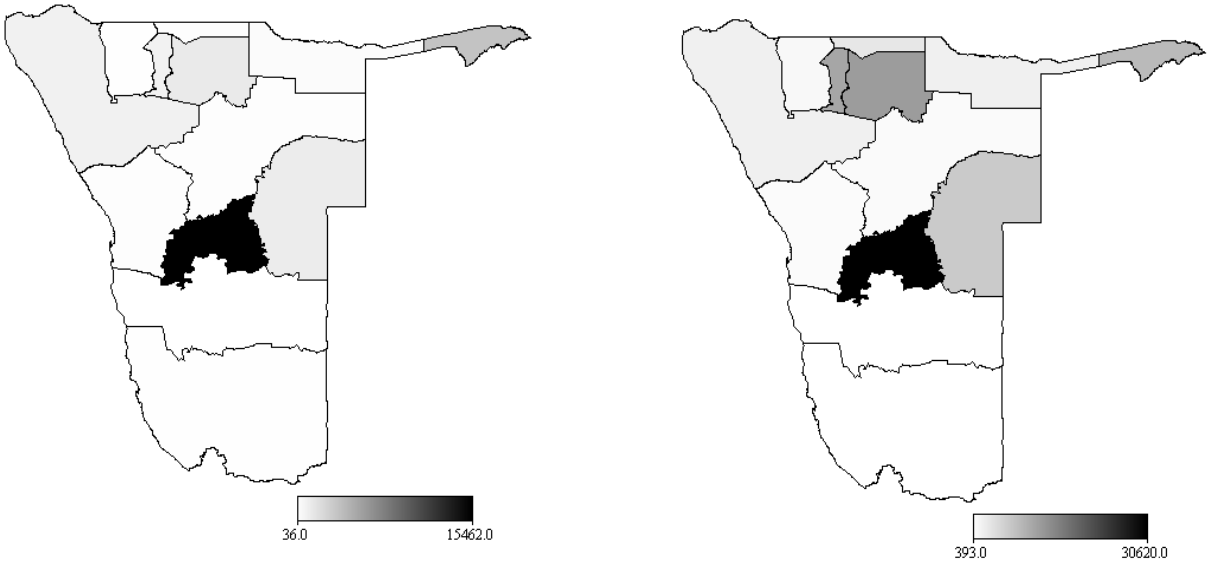


Figure 1: (a) Map of observed counts (total) of diabetes at regional level. (b) Map of observed counts (total) of hypertension at regional level. Given are the total observed counts at each region.

Kavango, Kunene and Ohangewena region also showed a relatively appeared low number of observed counts of hypertension as shown in Figure 1 (b).

4.3 Distribution of observed counts of diabetes and hypertension at health district level

Figure 2 (a) and Figure 2 (b) presents the distribution of diabetes and hypertension in 34 health districts in Namibia respectively. Health districts in darker grey colour indicate areas where diabetes and hypertension is high while the lighter grey colour show areas where diabetes and hypertension is low. The number of observed counts ranged from 39 to 15462 for diabetes while that of hypertension ranged from 393 to 30620.

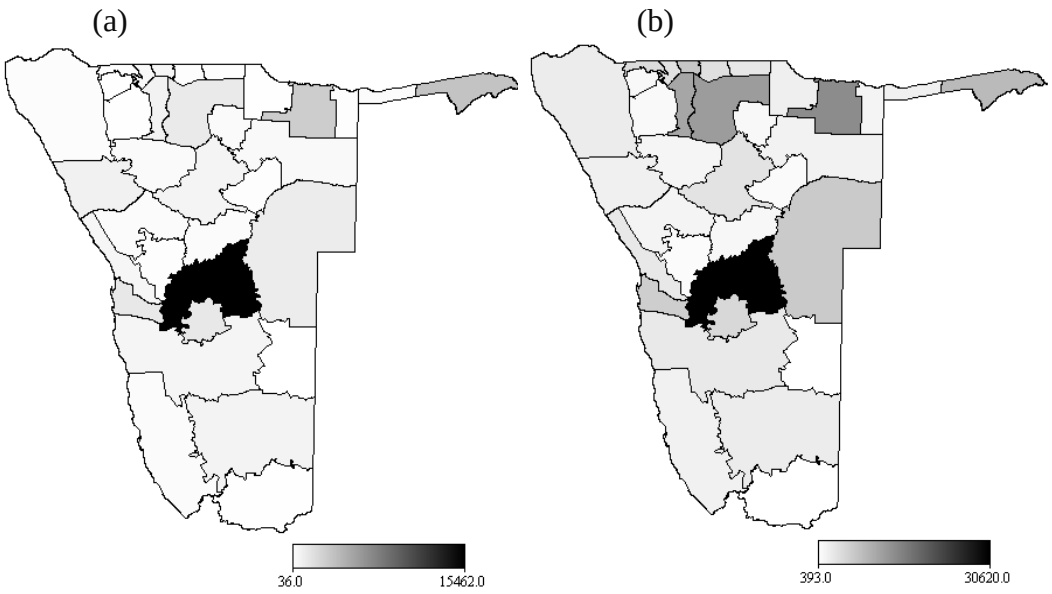


Figure 2: (a) Map of observed counts (total) of diabetes at health district level. (b) Map of observed counts (total) of hypertension at health district level. Given are total observed counts at each health district.

From Figure 2 (a), Windhoek health district had the highest observed counts of diabetes. Katima Mulilo health district and Rundu health district also showed that diabetes was slightly high in the health districts. The lowest counts of diabetes were found to be in Aranos health district. Nyangana, Nankundu health districts in Kavango region, Okahao, Outapi and Tsandi health districts in Omusati region were also found to have low observed counts of diabetes as shown in Figure 2 (a).

Figure 2 (b) also shows that hypertension varies across the health districts in Namibia with the highest observed counts found to be in Windhoek health district. Other areas where hypertension was high were Katima mulilo, Rundu, Onandjokwe and Oshakati

health districts. The lowest cases of hypertension were found in Aranos in Hardap region and Karasburg health district in Karas region.

4.4 Distribution of observed counts of diabetes and hypertension at constituency level

Figure 3 (a) and Figure 3 (b) presents the distribution of diabetes and hypertension counts in 107 constituencies in Namibia respectively. Constituencies with darker grey colours indicate areas where diabetes and hypertension is high while the lighter grey colours show areas where diabetes and hypertension is low.

The number of observed counts of diabetes ranged from 0 to 14215 with 0 representing no observed counts of diabetes in a constituency and 14215 representing the highest number of observed counts of diabetes in a constituency.

According to Figure 3 (a), there were not many constituencies found to have a high number of observed counts of diabetes as most constituencies exhibited a lighter grey colour. Khorixas, Rehoboth Rural appeared to have relatively high observed counts of diabetes but the highest appeared to be in Khomasdal North. Overall most constituencies had no darker grey colours suggesting a low number of observed counts in the different constituencies.

The distribution of observed counts hypertension is also given in Figure 3 (b). The observed counts of hypertension ranged from 0 to 22441 with 0 representing the minimum number of observed counts of hypertension in a constituency and 22441 representing the highest number of observed counts in a constituency.

From Figure 3 (b), there were not many constituencies found to have a high number of observed counts of hypertension. Rehoboth Rural, Gobabis appeared to have a high number of hypertension counts. Oniipa and Oshakati East also showed a high number of observed counts of hypertension. The highest number of observed counts was however found in Khomasdal North.

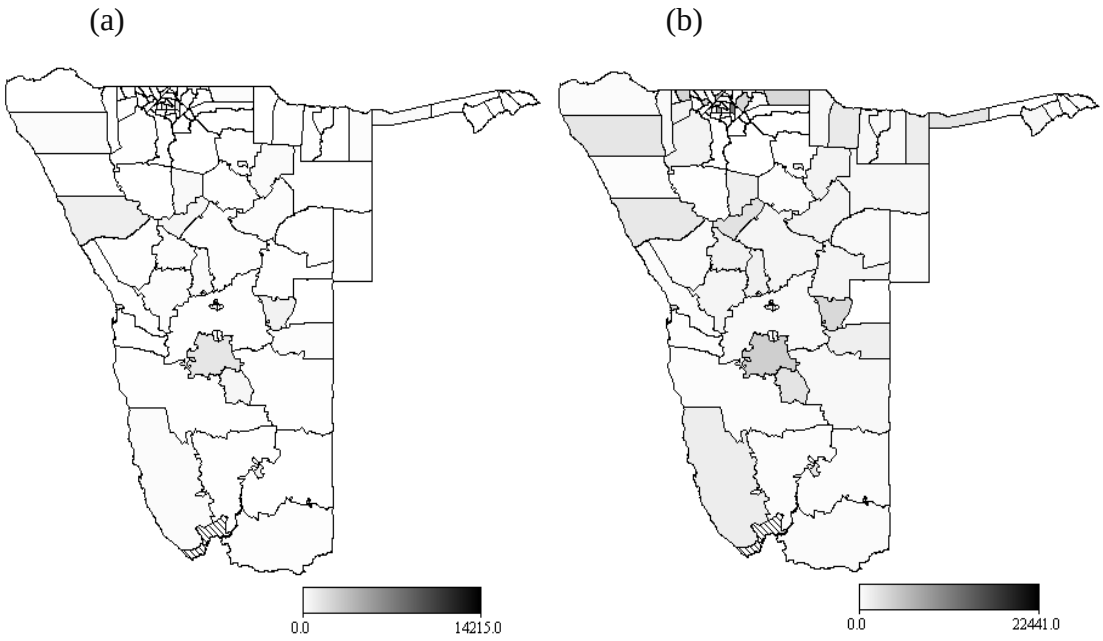


Figure 3: (a) Map of observed counts (total) of diabetes at constituency level. (b) Map of observed counts (total) of hypertension at constituency level. Given are the total observed counts at each constituency.

Overall, most constituencies had a low number of observed counts of diabetes as evident in Figure 3 (a). For example, Kahenge, Anamulenge, Oshikuku, Otjinene, Epukiro, Ruacana, Sesfontein, Arandis, Omaruru, Guinas and Berseba.

4.5 SMR for diabetes and hypertension at regional level

Standardized estimates of diabetes and hypertension were generated at regional level. The resulting maps of the SMR of diabetes and hypertension are given in Figure 4 (a) and Figure 4 (b) respectively. Darker grey colour represent areas where the SMR is high while areas that show lighter grey colours indicate areas where the SMR is low. The SMR ranged from 0.11 to 3.079 for diabetes while that of hypertension ranged from 0.31 to 5.097.

(a) (b)

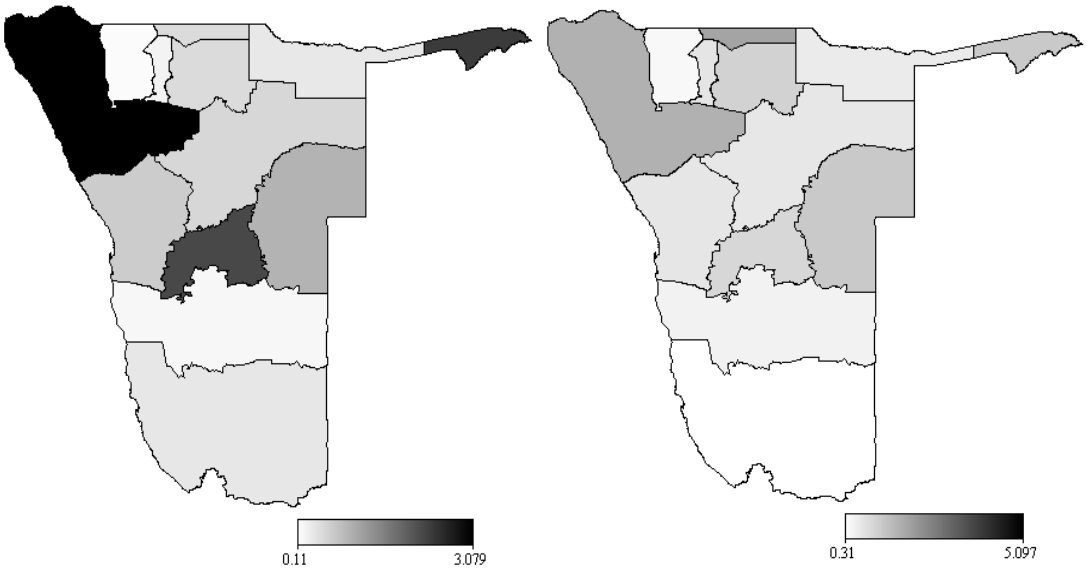


Figure 4: (a) Map of SMR of diabetes at regional level. (b) Map of SMR of hypertension at regional level. Given are the SMRs at each region.

The highest SMR for diabetes was found in Kunene region. Caprivi and Khomas region also had a high SMR as shown in Figure 4 (a). Omaheke region slightly had a darker grey colour suggesting a relatively high SMR of diabetes in the area. The lowest

SMR of diabetes was found in Hardap, Omusati and Oshana region. Other regions that seemed to exhibit low SMR of diabetes were Erongo, Karas, Kavango, Ohangwena and Otjozondjupa.

For hypertension, the SMR was highest in Kunene and Ohangwena region. Karas region had the lowest SMR of hypertension as shown in Figure 4 (b). Erongo, Hardap, Kavango, Omusati, Oshana and Otjozondjupa among others also exhibited a low SMR of hypertension.

4.6 SMR for diabetes and hypertension at health district level

Standardized estimates of diabetes and hypertension were generated to compare variation across sub regions. Darker grey colour indicate areas where the SMR is high while areas that show lighter grey colour indicate areas where the SMR is low. The results of the SMR for diabetes and hypertension at health district level are given in Figure 5 (a) and Figure 5 (b) respectively. The SMR for diabetes ranged from 0.11 to 3.079 while that of hypertension ranged from 0.31 to 5.097.

For diabetes, Khorixas health district was found to have the highest SMR compared to other health districts. Katima Mulilo, Nyangana, Rehoboth health districts also showed that the SMR was high in the areas as shown in Figure 5 (a). Slightly darker grey colors were evident in Keetmanshoop, Mariental, Omaheke, Outjo and Swakopmund among others, suggesting that the areas had a relatively high SMR of diabetes. Aranos and Nankundu had the lowest SMR while Eenhana, Engela, Oshikuku, Okahao, Oshakati,

Outapi and Tsandi health distrits had a relatively lower SMR as illustrated in Figure 5 (a).

For the SMR of hypertension, Nyangana health district had the highest SMR as shown in Figure 5 (b). Moreover, Okongo health district also exhibited a high SMR. Other areas that showed that the SMR was high include Khorixas, Nankundu and Rehoboth health districts. Katima mulilo, Onandjokwe, Omaheke and Omaruru health districts also had slightly high SMR as shown in Figure 5 (b). The lowest SMR was found in Karasburg health district.

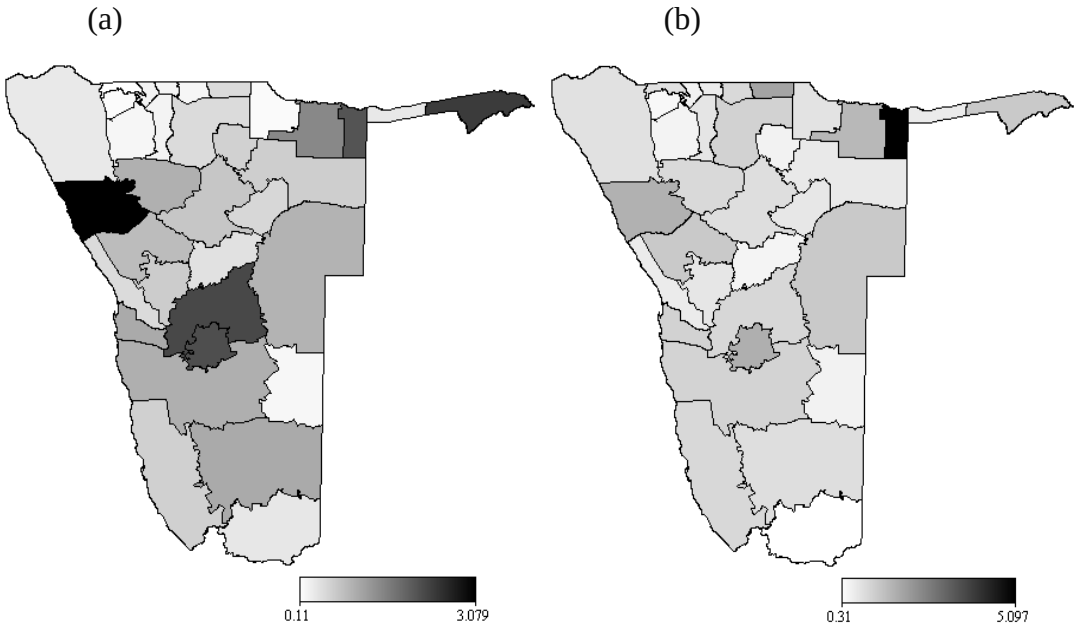


Figure 5: (a) Map of SMR of diabetes at health district level. (b) Map of SMR of hypertension at health district level. Given are SMRs at each health district.

4.7 SMR for diabetes and hypertension at constituency level

Figure 6 (a) and Figure 6 (b) presents the SMR estimates of diabetes and hypertension at constituency level. The SMR of diabetes at constituency level ranged from 0 to 15.102 while the SMR of hypertenison ranged from 0 to 3.019. Lighter grey colours indicate areas with a low SMR whereas constituencies that have a high SMR are indicated by a dark grey colour.

According to Figure 6 (a), the highest SMR of diabetes was found to be in Khomasdal North constituency in Khomas region. Khorixas and Rehoboth Rural constituency also showed a high SMR of diabetes as shown in Figure 6 (a). Most constituencies showed a low SMR as evident in Figure 6 with mostly light grey colours in those constituencies.

(a) (b)

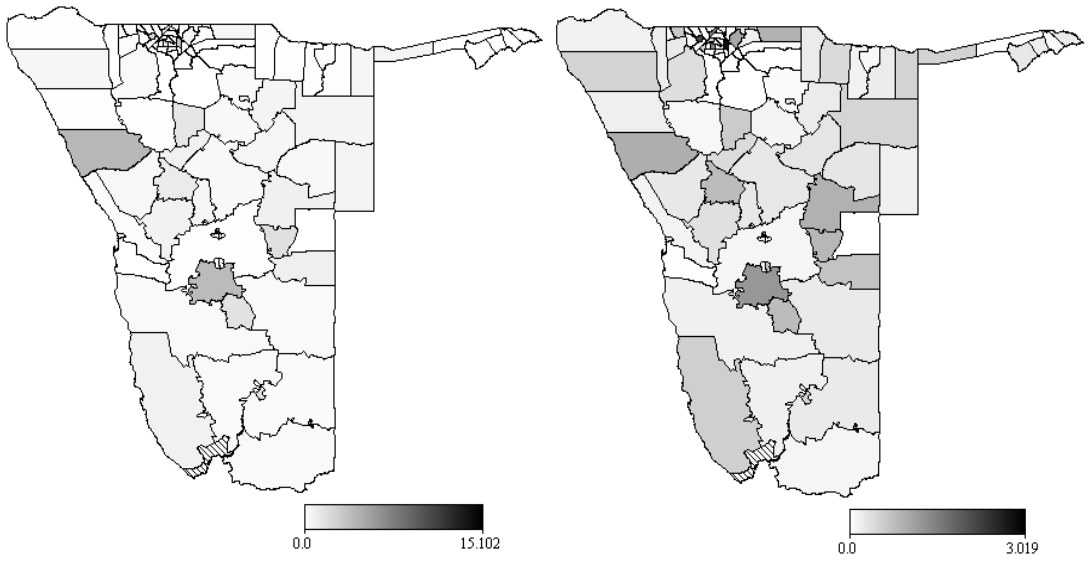


Figure 6: (a) SMR for diabetes at constituency level. (b) SMR for hypertension at constituency level. Given are the SMRs at each constituency.

A significant number of constituencies with a high SMR of hypertension was also found and is shown in Figure 6 (b). Constituencies such as Aminuis, Eenhana, Khorixas, Gobabis, Oniipa, Okongo, Oshikuku, Rehoboth Rural and Steinhausen all had a relatively high SMR of hypertension. The lowest SMR of hypertension were found in constituencies such as Ontamanzi, Omuntele,Uuvudhiya, Kalahari, Walvis Bay Rural and Rehoboth West Urban.

4.8 Spatial autocorrelation for diabetes and hypertension

4.8.1 Introduction

The spatially structured effects of diabetes and hypertension were mapped at region, health district and constituency level to give results of smoothed risk estimates from three BYM models fitted. A CAR distribution prior to account for over-dispersion and spatial autocorrelation for both diseases was used in modelling the RR of diabetes and hypertension.

4.8.2 Spatially structured effects of diabetes and hypertension at regional level

The RR of diabetes and hypertension at region level was mapped, the results are given in Figure 7 (a) and Figure 7 (b) respectively. The RR for diabetes ranged from 0.871 to 1.111 whereas RR that of hypertension ranged from 0.971 to 1.021. Areas with a high RR for diabetes and hypertension are represented on the map with darker grey colours while areas that have a low RR of diabetes and hypertension are shown with a lighter grey colour.

The RR of diabetes was found to be highest in Khomas and Karas region. Erongo, Hardap, Omaheke and Otjozondjupa region also showed a high RR of diabetes. Regions such as Caprivi, Kavango and Kunene showed a relatively high RR as shown in Figure 7 (a). The lowest RR of diabetes was evident in Oshana region. Ohangwena, Omusati and Oshikoto were also found to have a relatively low RR of diabetes.

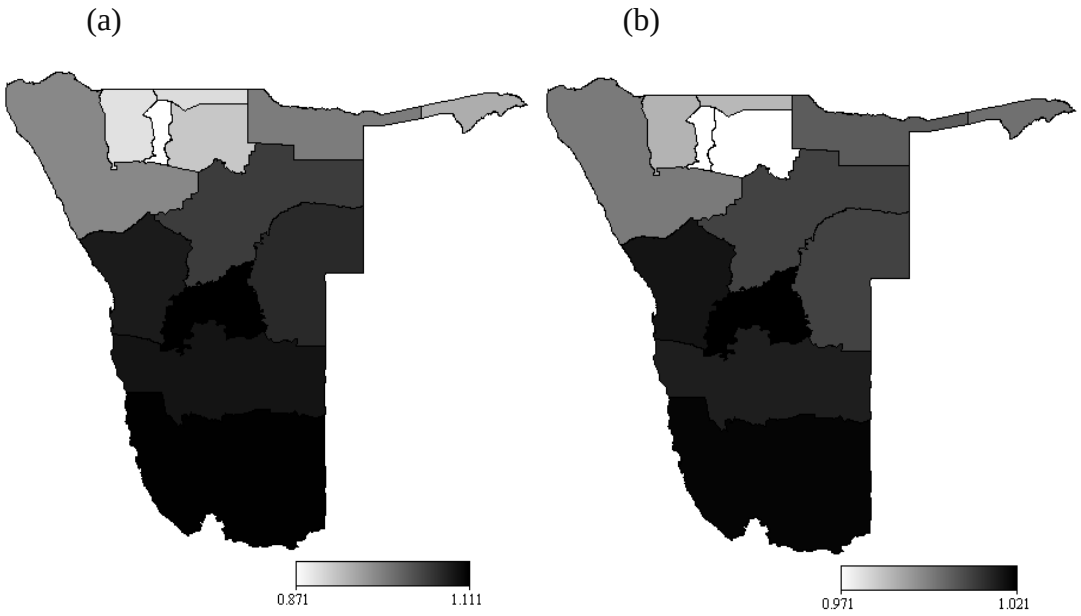


Figure 7: (a) Spatially structured effects of diabetes at region level. (b) Spatially structured effects of hypertension at region level. Given are the RRs at each region.

For hypertension at regional level, the highest RR was in Karas and Khomas region as shown in Figure 7 (b). Caprivi, Erongo, Hardap, Kavango, Kunene, Omaheke and Otjozondjupa region also showed a relatively high RR of hypertension. Figure 7 (b) also gives the results of regions with low RR of hypertension. The lowest RR of hypertension was found to be in Oshana and Oshikoto region. Ohangwena and Omusati region also appeared to show a low RR of hypertension.

4.8.3 Spatially structured effects of diabetes and hypertension at health district level

The RR of diabetes and hypertension at health district level were mapped, the RR is given in Figure 8 (a) and Figure 8 (b) respectively. Darker grey colours on the map represent areas where the RR of diabetes and hypertension are higher while areas with a low RR are represented as lighter grey colours. The RR for diabetes ranged from 0.886 to 1.085 while for hypertension, the RR ranged from 0.973 to 1.016.

(a) (b)

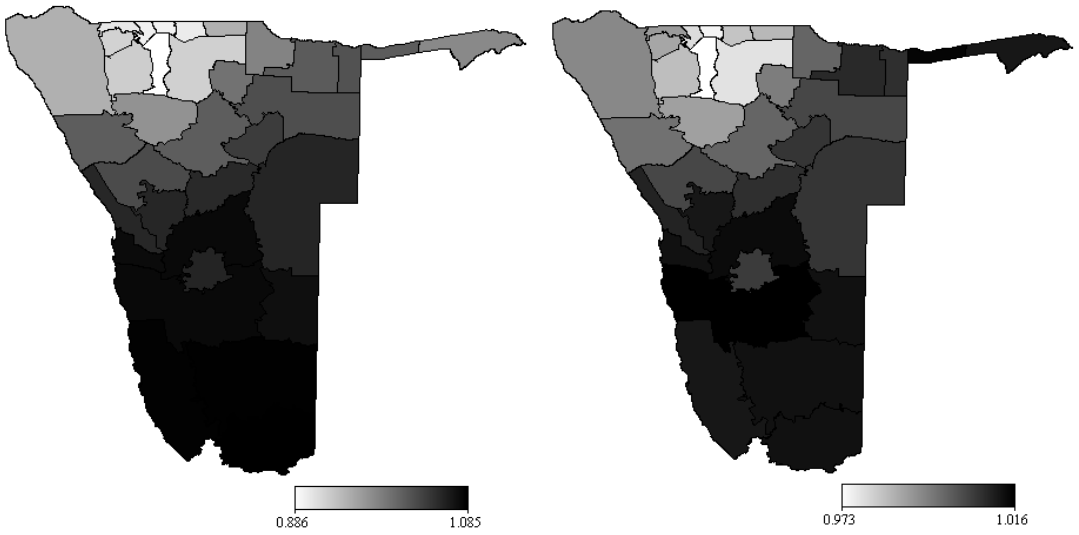


Figure 8: (a) Spatially structured effects of diabetes at health district level. (b) Spatially structured effects of hypertension at health district level. Given are the RRs at each health district.

The RR of diabetes was found to be highest in Karasburg, Keetmanshoop and Luderitz health district. Aranos, Katima Mulilo, Mariental, Walvis Bay, Windhoek, Rehoboth, Gobabis, Okahandja, Usakos, swakopmund, Okakarara, Grootfontein, Rundu and

Nyangana among others were also found to be health districts with a high RR of diabetes as shown in Figure 8 (a). The lowest RR of diabetes was found in Oshakati health district. Health districts such as Eenhana, Engela, Outapi, Oshikuku, Onandjokwe, Okahao and Tsandi also showed a relatively low RR of diabetes.

For hypertension, the highest RR was found in Andara and Mariental health districts as shown in Figure 8 (b). Health districts such as Aranos, Luderitz, Karasburg, Keetmashoop, Gobabis, Grootfontein, Okakarara, Okahandja, Omaruru, Nyangana, Nankudu, Rundu, Katima Mulilo, Opuwo, Khorixas, Swakopmund, Usakos, Walvis Bay and also showed a relatively high RR of hypertension. The lowest RR of hypertension was evident in Oshakati health district as shown in Figure 8 (b). Engela, Onandjokwe and Oshikuku health districts appeared to have a low RR of hypertension.

4.8.4 Spatially structured effects of diabetes and hypertension at constituency level

The results shown in Figure 9 (a) and Figure 9 (b) are the RR of diabetes and hypertension at constituency level respectively. For diabetes, the RR ranged from 0.904 to 1.123. The RR of hypertension ranged from 0.978 to 1.03. Areas in darker grey colours on the map represent areas with a high RR of diabetes and hypertension while areas that show a low RR are represented by a lighter grey colour.

The RR of diabetes showed significant spatial variation with areas with darker grey colours appearing mostly in the southern and central part of Namibia. The RR of

diabetes was high in the southern part of Namibia, specifically constituencies situated in the southern part of the region.

The RR of diabetes was mainly found to be high in constituencies such as Luderitz, Orangemund, Karasburg, Berseba, Gibeon, Keetmanshoop Rural, Marienatal Rural, Mariental Urban, Walvis Bay Rural, Rehoboth Rural, Aminuis, Gobabis, Windhoek Rural, Windhoek West, Windhoek East, Khomasdal North, Samora Machel, Tobias Hainyeko, Soweto, Moses Garoeb, Katutura East, Katutura Central and Rehoboth East Urban among others. Arandis, Karibib, Okahandja, Rundu Rural West, Steinhausen, Otjimbide also showed a relatively high RR of diabetes as shown in Figure 9 (a).

(a) (b)

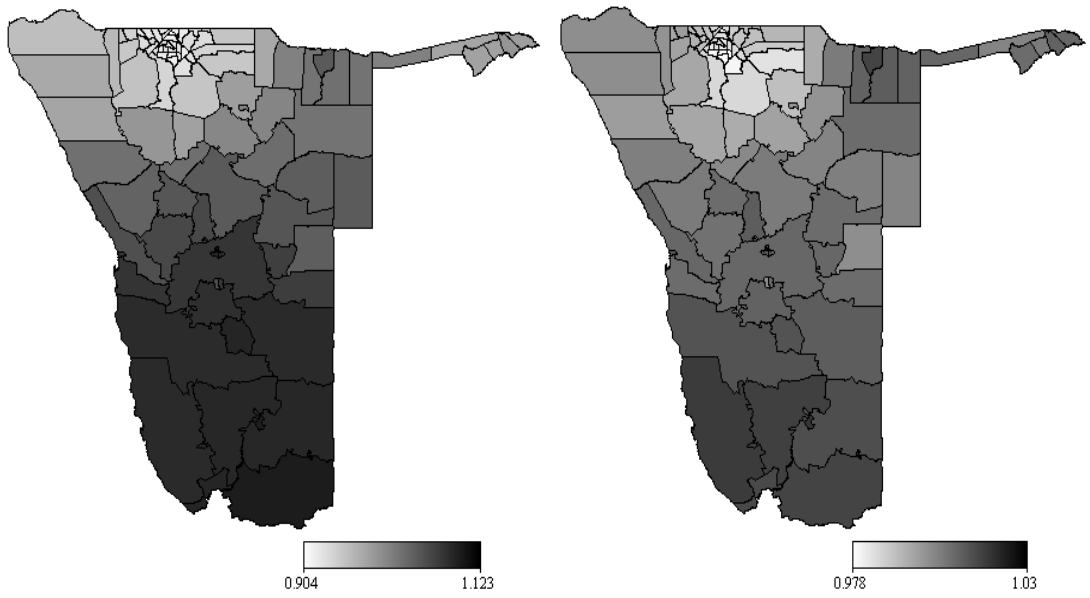


Figure 9: (a) Spatially structured effects of diabetes at constituency level. (b) Spatially structured effects of hypertension at constituency level. Given are the RRs at each constituency.

Areas with a low RR of diabetes were mostly in the northern part of Namibia. Figure 9 (a) shows that constituencies such Elim, Opundja, Ogongo, Okatana and Okatyali among others had a low RR of diabetes.

For the RR of hypertension at the constituency level, results from Figure 9 (b) shows that there was significant spatial variation. The RR was found to be highest among constituencies in the southern part of Namibia and lowest in the northern constituencies in Namibia. Some parts of the central part of Namibia had a reduced RR of hypertension. Some constituencies in the Kavango region and Caprivi region situated in the north eastern parts of Namibia had constituencies with a high RR whilst some of the constituencies had a low or reduced RR as shown in Figure 9 (b). Luderitz, Orangemund, Karasburg, Berseba, Gibeon, Keetmanshoop Rural, Marienatal Rural, Mariental Urban, Walvis Bay Rural, Rehoboth Rural, Aminuis, Gobabis, Windhoek Rural, Windhoek West, Windhoek East, Khomasdal North, Samora Machel, Tobias Hainyeko, Soweto, Moses Garoeb, Katutura East, Katutura Central and Rehoboth East Urban among others. Arandis, Karibib, Okahandja, Rundu Rural West, Steinhausen, Otjimbide also showed a relatively high RR of hypertension as shown in Figure 9 (b). Ndiyona, Mashare, Kapako, Kahenge, Rundu Rural East, Rundu Rural West, Mukwe, Kongola, Linyanti,Sibinda, katima Mulilo Rural and Kabe among others also showed a high RR of hypertension.

Low RR of hypertension were found to be in Omuntele, Okatyali, Ompundja,, Oshakati West and Oshakati East among others as shown in Figure 9 (b).

4.9 Multiscale modelling of diabetes and hypertension at regional, health district and constituency level

A total of twenty-two candidate models were fitted with the aim of generating RR maps at region, health district and constituency level. Eleven candidate models for each outcome variable were fitted (i.e. eleven models for diabetes and eleven models for hypertension). The models were fitted as follows:

- Fit a fixed effects model for each disease
- Fit a random effects model for each disease where health district only is considered as a spatial effect.
- Fit a random effects model for each disease where constituency only is considered as a spatial effect.
- Fit a random effects model for each disease where region only is considered as a spatial effect.
- Fit a random effects model for each disease with structured and unstructured random effects where health district only is considered as a spatial effect.
- Fit a random effects model for each disease with structured and unstructured random effects where constituency only is considered as a spatial effect.
- Fit a random effects model for each disease with structured and unstructured random effects where region only is considered as a spatial effect.

- Fit a multiscale random effects model without covariates for each disease by considering two different spatial effects into the model. The model fitted only considered spatially structured effects.
- Fit a multiscale random effects model without covariates for each disease by considering three different spatial effects into the model. The model fitted considered spatially structured and spatially unstructured effects.
- For each disease, observe the deviance, P_D, DIC and calculate the ΔDIC. The ΔDIC gives change in the deviance with respect to the deviance of the preceding model with -ΔDIC indicating an improved fit on the current model.
- Select any multiscale model using the ΔDIC rule of thumb that states that a ΔDIC less than 4 implies no significant difference between the models and therefore any model can be selected as a best fit model for the data.
- For each disease, add covariates to the multiscale random effects model selected based on the rule of thumb of ΔDIC. Again, for each disease, observe the deviance, P_D, DIC and ΔDIC and make inferences and conclusions based on the multiscale model with covariates if ΔDIC suggest an improved fit, otherwise make inferences and conclusions on multiscale model without covariates.

The study however begins by presenting the model estimates for deviance, P_D, DIC and ΔDIC of all models fitted (see Table 3 and Table 4). For DIC, only saturated DIC were reported.

4.9.1 Model Selection Criterion for diabetes

Table 3 shows results for models fitted for the outcome variable diabetes using BayesX. Deviance, P_D and DIC were each given from the output. The ΔDIC of all models was also calculated.

Model 1 (M_1) was a fixed effects model fitted first, model 2 (M_2), model 3 (M_3) and model 4 (M_4) were random effects models fitted and only considered spatially structured effects. Model 5 (M_5), model 6 (M_6) and model 7 (M_7) were models that considered spatially structured and spatially unstructured effects fitted at the same level (for example, M_5 only has health district as a spatial effect). Lastly, model 8 (M_8), model 9 (M_9), model 10 (M_{10}) and model 11 (M_{11}) were multiscale models fitted. M_8 and M_9 only considered spatially structured effects while M_{10} and M_{11} considered both spatially structured effects and spatially unstructured effects.

According to Table 3, M_1 was found to have a DIC equal to 64.1038. In addition, M_1 had the lowest DIC. Whilst M_1 had the lowest DIC, the model only had fixed effects and it was not reasonable to use the model for spatial variation of diabetes. The study therefore calculated the ΔDIC for all models as shown in Table 3.

The model with the highest DIC was found to be M_6 (DIC= 70.8387). M_{10} and M_{11} were also found to be have a high DIC (DIC of 70.1007 and 70.7490 respectively).

It was uncertain to decide on which model to select as the best model even though the fixed effects model M_1 had the lowest DIC. Moreover, the differences between the various models fitted showed that the ΔDIC was not large.

This allowed the present study to explore the use of the ΔDIC to select the model that would best explain the data. The rule of thumb on selecting the best model that explains the data suggests that ΔDIC < 4 implies that there is no significant difference between the models fitted and thus any model can be selected. The rule of thumb can also be relaxed to ΔDIC < 10. The current study however used ΔDIC < 4 with ΔDIC ≥ 4 implying significant differences between candidate models.

The study therefore selected M_{11} as the best model to explain spatial variation at different scales and in terms of epidemiology would be best suited to identify the risk at different levels considered for the study. According to Table 3, M_{11} had a DIC equal to 70.7490, the ΔDIC was found to be 0.6483 which was less than four. Moreover, the deviance of M_{11} was among found to have the lowest even though the penalization due to the number of effective parameters in the model was more as shown in Table 3. Furthermore, the deviance itself explains the goodness of fit of the model to the data.

Table 3: Deviance, PD, DIC and ΔDIC for models of diabetes fitted in BayesX

	Models for diabetes	Deviance	PD	DIC	ΔDIC
M_1:	diab=fixed effects only	52.5530	5.7754	64.1038	-
M_2:	diab=hd(spatial)	61.3641	3.5062	68.3764	4
M_3:	diab=const(spatial)	60.6957	4.2432	69.1822	0.8058
M_4:	diab=reg(spatial)	60.4191	3.7431	67.9053	-1.2769
M_5:	diab=hd(spatial)+hd(random)	59.1793	5.1928	69.5650	2

M_6:	diab=const(spatial)+const(random)	59.9705	5.4341	70.8387	1.2737
M_7:	diab=reg(spatial)+reg(random)	59.0619	4.8681	68.7981	-2.0406
M_8:	diab=hd(spatial)+const(spatial)	58.5720	5.1703	68.9126	0.1145
M_9:	diab=hd(spatial)+reg(spatial)	58.6599	4.8226	68.3051	-0.6075
M_{10}:	diab=hd(spatial)+hd(random))+const(spatial)	57.3224	6.3891	70.1007	1.7956
M_{11}:	diab=const(spatial)+const(random))+hd(spatial)+reg(spatial)	56.0261	7.3614	70.7490	0.6483
M_{12}:	**diab=const(spatial)+const(random))+hd(spatial)+reg(spatial)+ covariates**	**49.2165**	**10.0710**	**69.3585**	**-1.3905**

Note: diab refers to diabetes, hd refers to health district, const refers to constituency and reg refers to region.*

Covariates were then added to the multiscale model (M_{11}) and the deviance, P_D, DIC and ΔDIC were observed. The results of the multiscale model with covariates fitted [model 12 (M_{12}) are also given in Table 3. The study found the DIC of M_{12} to be 69.3585 suggesting that adding covariates to M_{11} produced an improved model fit. The deviance of M_{12} was also found to be quite small as compared to M_{11} and other models fitted for diabetes as shown in Table 3. The ΔDIC of M_{12} was also less than four ($\Delta\text{DIC} = -1.3905$), suggesting that the model was not distinguishable from M_{11}. The study therefore selected M_{12} as the best model to explain spatial variation at

region, health district and constituency level. The results of diabetes were therefore interpreted based on the multiscale model with covariates fitted.

4.9.2 Model selection criterion for hypertension

A similar approach used in fitting models for diabetes was used to fit and select the best model to examine the spatial variation of the outcome variable hypertension. Table 4 shows results for models fitted using BayesX. The deviance, P_D and DIC were each given from the output. The ΔDIC of all models was also calculated.

Model 13 (M_{13}) was a fixed effects model fitted first, model 14 (M_{14}), model 15 (M_{15}) and model 16 (M_{16}) were random effects models fitted and only considered spatially structured effects. Model 17 (M_{17}), model 18 (M_{18}) and model 19 (M_{19}) were models that considered spatially structured and spatially unstructured effects fitted at the same level (for example, M_{19} only has region as a spatial effect). Lastly, model 20 (M_{20}), model 21 (M_{21}), model 22 (M_{22}) and model 23 (M_{23}) were multiscale models fitted. M_{20} and M_{21} only considered spatially structured effects while M_{22} and M_{23} considered both spatially structured effects and spatially unstructured effects.

Table 4: Deviance, PD, DIC and ΔDIC for models of hypertension fitted in BayesX

	Models for diabetes	Deviance	PD	DIC	ΔDIC
M_{13}:	ht=fixed effects only	43.1411	5.8104	54.7619	-
M_{14}:	ht=hd(spatial)	46.7582; pp	2.1880	51.1341	-3.6278
M_{15}:	ht=const(spatial)	46.5086	2.3343	51.1772	0.0431

M_{16}:	ht=reg(spatial)	46.4885	2.3234	51.1353	-0.0419
M_{17}:	ht=hd(spatial)+hd(random)	45.4816	3.6846	52.8508	1.7155
M_{18}:	ht=const(spatial)+const(random)	45.7223	3.7541	53.2305	0.3797
M_{19}:	ht=reg(spatial)+reg(random)	45.1638	3.4556	52.0750	-1.1555
M_{20}:	ht=hd(spatial)+const(spatial)	45.9027	3.2216	52.5459	0.4709
M_{21}:	ht=hd(spatial)+reg(spatial)	45.8502	3.1994	52.2490	-0.2969
M_{22}:	ht=hd(spatial)+hd(random))+const(spatial)	44.8702	4.7050	54.2803	2.0313
M_{23}:	**ht=const(spatial)+const(random))+hd(spatial)+reg(spatial)**	**44.2738**	**5.6490**	**55.5718**	**1.2915**
M_{24}:	ht=const(spatial)+const(random))+hd(spatial)+reg(spatial)+ covariates	40.3813	10.3765	61.1343	5.5625

Note: ht refers to hypertension, hd refers to health district, const refers to constituency and reg refers to region.*

According to Table 4, M_{14} was found to have the lowest DIC (DIC= 51.1341) but only had health district as a spatial effect implying that the study could not fully evaluate the effects of region and constituency if the model was selected. It was therefore reasonable to calculate the ΔDIC and select the best fit model based on the rule of thumb suggesting that ΔDIC less than four implies no significant differences between the candidate models and therefore any model could be selected as the best fit model that explains the data. Moreover, the values of DIC of the different models fitted for

hypertension seemed indistinguishable from one others or small differences of DIC between the candidate models were observed as shown in Table 4.

The model with the highest DIC was found to be M_{23} (DIC=55.5718). Similar models with found to have a high DIC were models M_{18} and M_{22}(DIC equals to 53.2305 and 54.2803 respectively). Even though M_{23} had the highest ΔDIC compared to other candidate models, the ΔDIC was less than four (ΔDIC= 1.2915) suggesting the model could be used to explain the spatial variation of hypertension at different level considered for the study. The differences between the various models fitted showed that the ΔDIC was not large.

The study therefore applied the rule of thumb on selecting the best model that explains the data which entails that $\Delta\text{DIC} < 4$ implies that there is no significant difference between the models fitted and thus any model can be selected.

The study therefore selected M_{23} as the best model to explain spatial variation because the model was more suited to provide results at different levels which is an important component in disease mapping. From Table 4, the study found the deviance of M_{23} to be 44.2738 which was among the lowest among the different candidate models. The deviance measures the goodness of fit of the different models to the data, again from Table 4, the deviance values of the different models were found to have small differences from one another.

Covariates were then added to the multiscale model (M_{23}) and the deviance, P_D, DIC and ΔDIC were observed. The results of the multiscale model with covariates fitted [model 24 (M_{24})] are also given in Table 4.

The study found the DIC of M_{24} to be 61.1343 suggesting that adding covariates to M_{23} produced a poor fit model. The results from $M_{24}$was not surprising because much of the variation had already been explained in M_{23} because considering health district, region or constituency as spatial effects and then adding covariates produced a poor fit model. This could be because adding health district as a spatial effect is failing because much of the variation captured by the health districts is already explained under region as a spatial effect, therefore adding the health district component proved not to improve the model. In addition, much of the variation of constituency as a spatial effect might have been already explained under region as a spatial effect.

The study also found the deviance of M_{24} to be quite small as compared to M_{23} and other models fitted for hypertension as shown in Table 4. The ΔDIC of M_{24} was however found to be also greater than four ($\Delta\text{DIC} = 5.5625$), suggesting that the model was distinguishable from M_{23}. The decision rule for explaining spatial variation of hypertension at health district, region and constituency level was to use M_{23}, a multiscale model without covariates. The results of hypertension were therefore interpreted based on the multiscale model without covariates fitted.

4.9.3 Residual plots for fitted models

Residual plots were used to assess model goodness of fit. Figure 10 (a) and Figure 10 (b) shows the results of residual boxplots for all models fitted. The closer the median of the boxplot is to the mean zero, the better the model fits the data. In addition, the narrower the boxplot, the more the model explains much of the variation suggesting that residuals are small which might agree with the DIC for the best model. Residual boxplots for the models of diabetes are given Figure 10 (a) while those of hypertension are given in Figure 10 (b).

From Figure 10 (a), most of the boxplots were closer to the mean zero, for instance M_4, M_5, M_7, M_8, M_9, M_{10}, M_{11} and M_{12}. For the multiscale models M_8, M_9, M_{10}, M_{11} and M_{12}, the residual plots in Figure 10 (a) also show that the median is closer to the mean zero suggesting that the models are also a better fit for the data. Moreover, M_2, M_3, M_{10}, M_{11} and M_{12} seem to exhibit narrower boxplots implying that the models explains more variation than the other models. In addition because the boxplot of residuals for M_2, M_3, M_{10}, M_{11} and M_{12} appear narrower, it suggests that the residuals are small.

For hypertension, the median of most residual boxplots were also found to be closer to the mean zero as shown in Figure 10 (b) suggesting that the fitted models seemed to fit the data well. For instance, M_{14}, M_{15}, M_{17}, M_{18}, M_{19}, M_{21} and M_{23} all appeared to be closer to the mean zero. The residual boxplot of M_{24} is given in Figure 10 (b), the median of the boxplot is also closer to the mean zero suggesting that the model might fit the data well.

(a) (b)

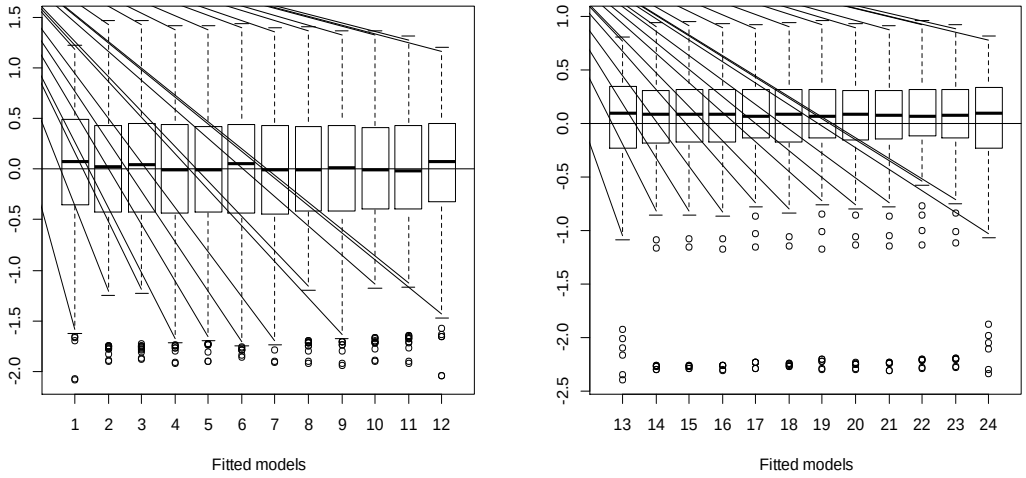


Figure 10: (a) Residual boxplots of fitted models for diabetes. (b) Residual boxplots of fitted models for hypertension.

Residual boxplots of M_{14}, M_{15}, M_{17}, M_{18}, M_{19}, M_{21} and M_{23} also appeared to exhibit some narrowness among the rest of the residual boxplots suggesting that the model does could explain more variation than the other models fitted.

4.9.4 Model estimates

The structured and unstructured spatial effects at health district, region and constituency level were mapped to come up with the RR of diabetes and hypertension at region, health district and constituency level.

4.9.4.1 Structured spatial variation of diabetes and hypertension at regional level

To capture the spatial variation of diabetes and hypertension at regional level, the RR for diabetes and hypertension was mapped using the multiscale models fitted (M_{12} and

M_{23}), results of the maps of RR for diabetes and hypertension are given in Figure 11 (a) and Figure 11 (b) respectively. For diabetes, the RR ranged from 0.969 to 1.027 while for hypertension the RR ranged from 0.976 to 1.019. Darker grey colours on the maps indicate areas where the RR is high while areas with light grey colours indicate areas where the RR of diabetes and hypertension is low.

(a) (b)

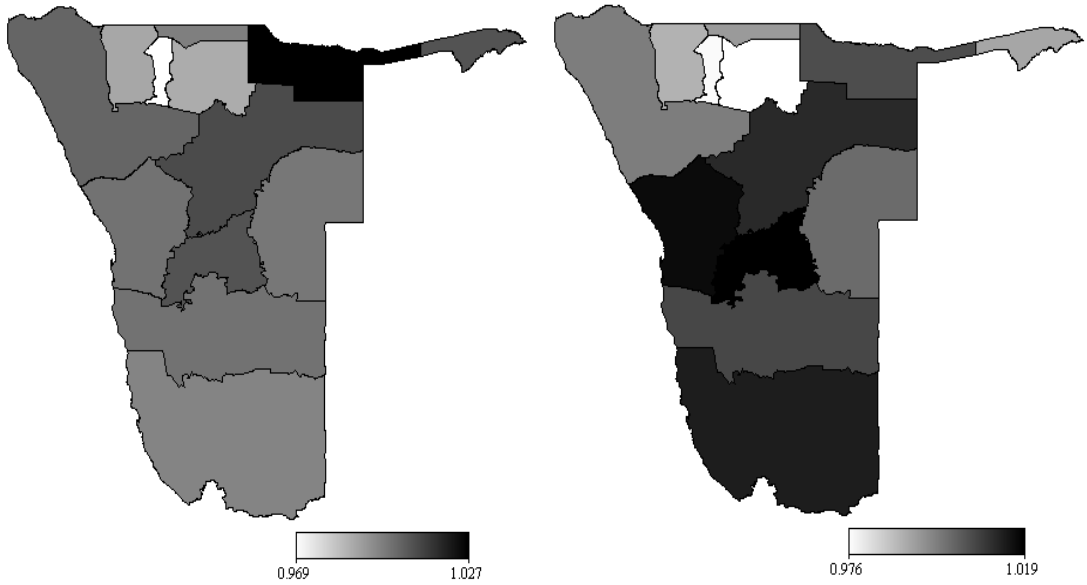


Figure 11: (a) Spatial variation in diabetes at regional level. (b) Spatial variation in hypertension at regional level. Given are RRs at each region.

From the map in Figure 11 (a), Kavango region had the highest RR for diabetes. Caprivi, Khomas and Otjozondjupa region also showed a relatively high RR of diabetes. Other regions with a somewhat high RR of diabetes were Erongo, Hardap, Karas, Kunene, Ohangwena and Omaheke. The lowest RR of diabetes was found in Oshana region as shown in Figure 11 (a). Other regions with low RR were Omusati and Oshikoto with light grey colours.

The map of RR of hypertension at regional level is also shown in Figure 11 (b). Khomas region was found to have the highest RR of hypertension. Erongo, Hardap Otjozondupa, Kavango and Karas region also appeared to exhibit a high RR of hypertension. Caprivi, Ohangwena, Kunene, Omaheke and Omusati was found to have a somewhat relatively high RR with Omusati exhibiting a reduced RR at regional level. The region with the lowest RR of hypertension was Oshikoto. Oshana region also had a low RR of hypertension.

4.9.4.2 Structured spatial variation of diabetes and hypertension at health district

Figure 12 (a) and Figure 12 (b) presents the structured spatial variation of diabetes and hypertension estimated from the multiscale models (M_{12} and M_{23}) at health district level. Figure 12 (a) shows that the RR for diabetes ranged from 0.967 to 1.037, while that of hypertension ranged from 0.975 to 1.012 as shown in Figure 12 (b). Darker colours on the maps represent areas where the RR for diabetes and hypertension is high while light grey colours on the map indicate areas where the RR of diabetes and hypertension is relatively low.

The highest RR of diabetes was found in Rundu health district as shown in Figure 12 (a). Andara, Nyangana, Nankudu, Grootfontein and Okakarara health districts also showed a high RR of diabetes. Aranos, Omaheke, Okahandja, Otjiwarongo, Omaruru, Khorixas, Kongo and Opuwo were also found to have a relatively high RR of diabetes. Karasburg, Luderitz and Windhoek health districts had a reduced RR.

There was a significant number of health districts with a low RR of diabetes. The lowest RR among health districts was found to be in Oshakati health district in Oshana

region. Other health district with low RR included Eenhana, Engela and Oshikuku and among others.

Figure 12 (b) also gives results of the RR of hypertension, the results showed that spatial variation exist amongst the different health districts in Namibia. The highest RR of hypertension was found in Andara, Rundu, Swakopmund and Windhoek health districts. There was also a relatively high RR of hypertension observed in some health districts. Health districts such as Aranos, Katima Mulilo, Luderitz, Okahandja and Usakos. Karasburg, Keetmashoop, Mariental, Gobabis, Omaruru, Khorixas, Okakarara, Grootfontein, Nyangana, Outapi and Nankudu also appeared to have a relatively high RR of hypertension.

(a) (b)

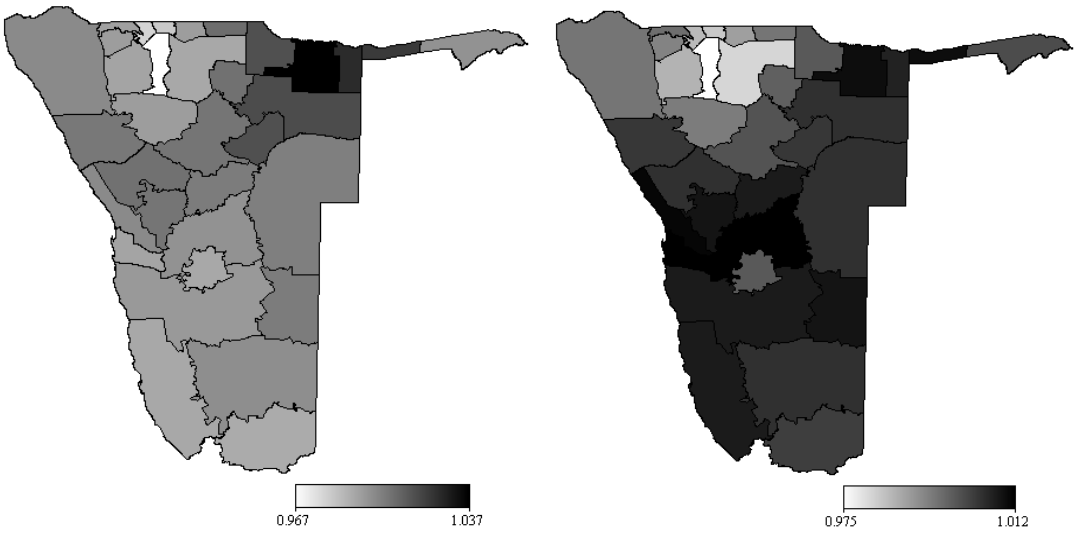


Figure 12: (a) Spatial variation in diabetes at health district level. (b) Spatial variation in hypertension at health district level. Given are RRs at each health district.

The lowest RR of hypertension was evident in Oshakati health district. Other health districts that exhibited a relatively low RR of hypertension were Engela, Onandjokwe, Okahao and Oshikuku.

4.9.4.3 Structured spatial variation of diabetes and hypertension at constituency level

The RR of diabetes and hypertension at constituency level is given in Figure 13 (a) and Figure 13 (b) respectively. The RR ranges of diabetes ranged from 0.972 to 1.064 while that of hypertension ranged from 0.967 to 1.03. Darker grey colours on the map indicate constituencies with a high RR while areas with a lighter grey colour represent areas with a low RR.

(a) (b)

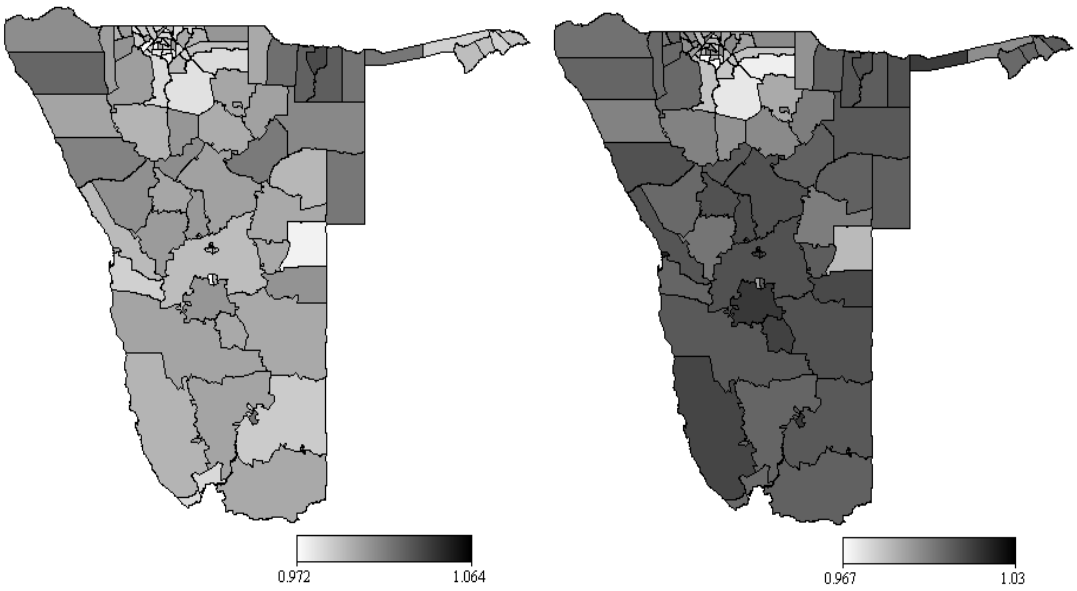


Figure 13: (a) Spatial variation in diabetes at constituency level. (b) Spatial variation in hypertension at constituency level. Given are RRs at each constituency.

Rundu Urban was found to have the highest RR of diabetes at constituency level. Other that were found to have a high RR of diabetes at constituency level were Mashare,

Kapako, Opuwo, Kahenge, Katima Mulilo Urban, Keetmanshoop Urban, Ndiyona, Khomasdal North, Mukwe, Otjimbinde, Okakarara, Khorixas, Tsumkwe, tsandi, Otjiwarongo, Epupa, Aminuis, Okongo, Outjo, Katutura Central, Rehoboth East Urban, Walvis Bay Urban and Daures. Constituencies such as Oniipa, Okahandja, Omatako, Grootfontein, Ruacana, Okahao, Outapi, Engela, Sesfontein, Tobias Hainyeko, Berseba, Gibeon, Karibib and Tsumeb among others.

The study found the lowest RR of diabetes in Rehoboth West Urban constituency. Constituencies such as Oranjemund, Keetmanshoop Rural, Kalahari, Uuvudhiya, Omuthiya, Eengondi, Elim, Anamulenge, Ogongo, Okaku, Okatyali, Opundja, Uukwiyu, Onayena and Okankolo among others had a low RR of diabetes as shown in Figure 13 (a).

With respect to hypertension, Rundu Urban was found to have the highest RR of hypertension at constituency level. Other constituencies that appeared to have a high RR were Katima Mulilo Urban, Khomasdal North, Walvis Bay Urban, Swakopmund, Katutura Central, Rehoboth East Urban, Keetmanshoop Urban, Mukwe, Katutura East, Mariental Urban, Tobias Hainyeko, Windhoek West, Luderitz, Windhoek East, Aminuis, Samora Machel, Okahandja, Omaruru, Mariental Rural, Ndiyona, Rundu Rural West, Windhoek Rural, Khorixas, Omatako, Arandis, Kapako, Gibeon, Keetmashoop Rural, Otjiwarongo, Tsumkwe, Kabe, Karasburg, Gobabis and Okakarara among others.

Constituencies such as Ruacana, Tsandi, Epupa, Karibib, Linyanti and Katima Mulilo Rural among others were found to have a reduced RR of hypertension as shown in Figure 13 (b). The lowest RR of hypertension was however found in Ompundja

constituency. Constituencies such as Okatyali, Eengondi, Omuthiya gwiimpundi, Olukonda, Ohangwena, Okaku, Omuntele, Onyaanya, Onayena, Okankolo, Uuvudhiya, Uukwiyu, Etayi, Kalahari, Oshakati East, Okatana, Endola, Guinas and Ondobe also appeared to exhibit a low RR of hypertension.

4.9.4.4 Fixed effects for best model of diabetes

The covariates considered for the study were obtained from the 2011 housing and population census of Namibia. The estimates for the fixed effects of the best model of diabetes (M_{12}) are presented in Table 5. The fixed effects for the best model of hypertension (M_{23}) were not estimated since the best model had no covariates.

The posterior means (given as RR) given in Table 5 fall within the 95% credible interval that is also reported in Table 5. For a fixed effect to be statistically significant, its RR should be greater than one and its corresponding 95 % credible interval should also be greater than one suggesting a statistically significant association.

Table 5: Estimates of fixed effects for the best model of diabetes. Given are the posterior means (RR) and 95 % credible intervals in brackets.

Models	**RR (95 % credible interval)**
Covariates	
Constant	1.5805 (1.3498; 1.8527)
Safe water	1.0046 (0.9905; 1.0191)
Wood/Charcoal	1.0051 (0.9945; 1.0160)

Main source of income: Wages and salaries	1.0104 (0.9961; 1.0249)
Main source of income: Pension	1.0007 (0.9812; 1.0219)
Education attainment: Incomplete primary	0.9901 (0.9653; 1.0129)

Note: *Description of covariates and models fitted are provided in table 1 and 2 respectively.*

In addition, a fixed effect is also statistically significant if its RR is less than one with the corresponding 95 % credible intervals also be less taking on values less than one. Moreover, for an increased risk for diabetes, the RR should be greater than one and the corresponding lower and upper intervals should also greater than one. For a reduced risk of diabetes, the RR and the lower and upper 95 % credible interval of a fixed effect should be less than one.

The RR for the fixed effects safe water, wood/charcoal, source of income: salaries and wages and source of income: pension were found to be positive and greater than one. In addition, the corresponding 95 % lower credible intervals were all less than one while the 95 % upper credible intervals were greater than one implying that the fixed effects were not statistically significant as shown in Table 5.

With respect to education attainment: incomplete primary education, the RR (0.9901) was less than one and the 95 % lower credible interval (0.9653) was also found to be

less than one. Furthermore, the 95 % upper credible interval (1.0129) was also found to be greater than one. Taking the RR and corresponding 95 % credible interval of the fixed effect Education attainment: incomplete primary, there was no statistically significant association found between diabetes and education attainment: incomplete primary.

4.9.4.5 Variance components of the best model for diabetes and hypertension

Table 6 presents the posterior means and 95% credible intervals of the variance components for the multiscale models M_{12} and M_{23}. The variance component is important for the study as it provides information on how large the RR is in a particular area.

Table 6: Variance components of the best model of diabetes and hypertension. Given are the posterior means and 95% credible intervals in brackets.

Disease	**Variance component**	**Posterior mean(95 % credible interval)**
Diabetes(M_{12})	Region	1.0262 (1.0005; 1.1705)
	Health district	1.0236 (1.0007; 1.1602)
	Constituency	1.0155 (1.0006; 1.0882)
Hypertension(M_{23})	Region	1.0168 (1.0006; 1.1069)
	Health district	1.0125 (1.0005; 1.0702)
	Constituency	1.0095 (1.0006; 1.0491)

For diabetes, the study found region as the variance component with a high posterior mean (1.0262) for model M_{12} that considered health district, region and constituency as spatial effects.

With respect to health district and constituency, the study found a posterior mean of 1.0236 and 1.0155 respectively. The high posterior mean of region compared to the spatial effects health district and constituency suggests that the RR for diabetes was large among the different regions and smaller in health districts and constituencies.

With respect to hypertension, the study also found region to have the highest variance component (posterior mean= 1.0168) as shown in Table 6. For health district and constituency, the posterior mean was 1.0125 and 1.0095 respectively. The high posterior mean for region compared to spatial effects health district and constituency considered in M_{23} suggests that the risk for hypertension was high in different regions compared to health districts and constituencies.

4.9.4.6 Spatially unstructured random effects for diabetes and hypertension at constituency level

The study presented the spatially unstructured random effects for diabetes and hypertension at constituency level, the results are shown in Figure 14 (a) and Figure 14 (b) respectively.

The RR for diabetes ranged from 0.98 to 1.017 while that of hypertension ranged from 0.985 to 1.011. Darker grey colours on the map represent areas with high RR of

diabetes and hypertension while areas with light grey colours indicate that the areas have a low RR of diabetes and hypertension.

(a) (b)

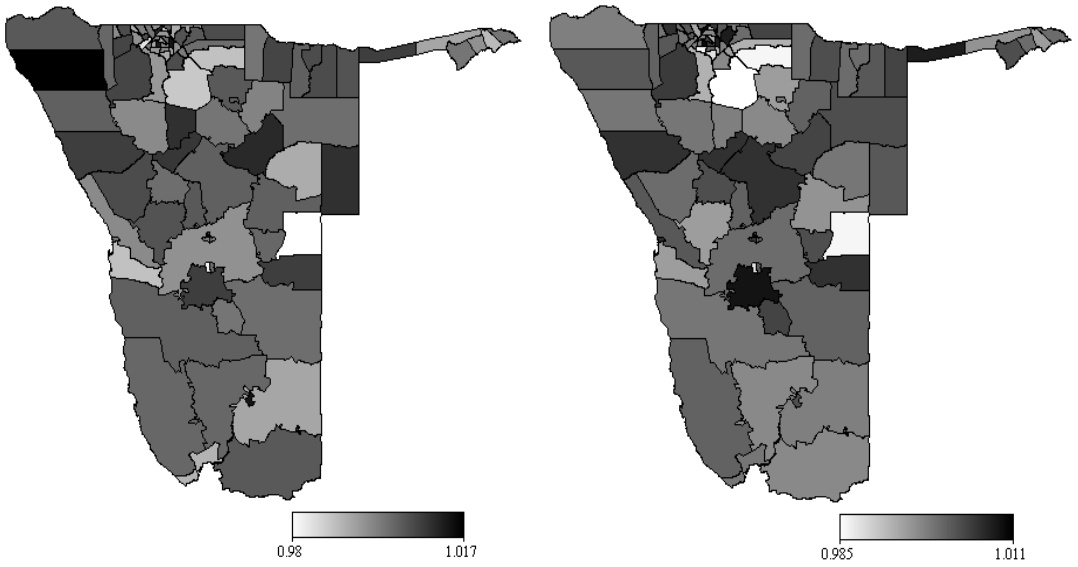


Figure 14: (a) Spatially unstructured effects of diabetes at constituency level. (b) Spatially unstructured effects of hypertension at constituency level. Given are RRs at each constituency.

Opuwo constituency was found to have the highest RR of diabetes at constituency level. Other constituecnies such as Khomasdal North, Keetmasnhoop Urban, Katima Mulilo Urban, Oniipa, Oshakati East, Rundu Urban, Engela, Okakarara, Outjo, Otjombinde, Otjiwarongo, Rehoboth West Urban, Mukwe, Khorixas, Aminuis, Okahao, Tsandi, Daures, Mashare, Rundu Rural West, Okongo, Oshikuku, Outapi, Omuntele, Karibib, Walvis Bay Urban, Ndiyona, Rundu Rural West, Tobias Hainyeko and Eenhana were also found to have a high RR of diabetes as shown in Figure 14(a).

A reduced RR of diabetes at constituency level was also found in constituencies such as Karasburg, Katutura central, Epupa, Epembe, Guinas, Okahandja, Gibeon, Kapako, Oshikango, Steinhausen, Tsumeb, Ruacana, Gobabis and Berseba.

The RR of diabetes was found to be lowest in Kalahari constituency as shown in Figure 14 (a). Other constituencies that showed a low RR of diabetes at constituency level include Rehoboth West Urban, Elim, Omuthiyagwiipundi, eengondi, Ohangwena, Walvis Bay Rural, Katima Mulilo Rural, Oranjemund, Ongwediva, Okatyali, Anamulenge, Otjinene, Ondobe, Oshakati West, Keetmashoop Rural, Olukonda, Uuvudhiya, Ompundja, Samora Machel, Endola and Onayena among others.

With respect to Hypertension, Figure 14 (b) shows that the RR was highest in Khomasdal North constituency. Other constituencies that appeared to have a high RR of hypertension were Katima Mulilo Urban, Engela, Oshakati East, Rehoboth East Urban, Mukwe, Eenhana, Oniipa, Rundu Urban, Swakopmund, Walvis Bay Urban, Khorixas, Aminuis, Outapi, Omatako, Otjiwarongo, Windhoek West, Oshikango, Okahao and Ondangwa.

A reduced RR of hypertension was also found in Ongwediva, Okakarara, Linyanti, Keetmashoop Urban, Katutura Central, Kahenge and Arandis among others. The lowest RR of hypertension was however found in Omuthiya Qwiimpundi constituency. Constituencies such as Eengondi, Ompundja, Kalahari, Okatyali, Rehoboth West Urban, Ohangwena, Olukonda, Okankolo, Uuvudhiya, Kongola and Otavi Constituencies among others also showed a low RR of hypertension.

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1 Discussion

The rising trend in NCDs calls for interventions that will assist in reducing the incidence and prevalence of diabetes and hypertension. In the same vein, while many health funding agencies and policy makers are reluctant to divert scarce resources away from communicable diseases into other areas of disease burden such as NCDs (Kandala, Tigbe, Manda & Stranges, 2013). The failure to divert these scarce resources to NCDs to address the problems associated with NCDs may impose significant burden for health sectors and the economy of Sub Saharan African countries (including Namibia) because few people would seek treatment, which could lead to high morbidity and mortality rates from potentially preventable diseases (Belue et al., 2009).

Although, government and other relevant sectors such as civil society, faith based organisations, UN agencies, and individuals amongst others have made considerable efforts and spend resources on measures pertaining to the reduction of NCDs in the Namibia, less efforts on understanding the contributing factors to the increase in the RR of NCDs has been achieved. For example, the study found limited literature on studies that have attempted to model diabetes and hypertension at region, health district and constituency level. Furthermore, of the studies found, the modeling of disease was primary done at one level of analysis. In the same vein, these studies have mostly attempted to model communicable diseases.

The study aimed to apply a multiscale model for diabetes and hypertension to identify spatial variations of diabetes and hypertension at various geographic levels in Namibia to guide better planning, monitoring and evaluation and assist in targeting of resources. Specific objectives for the study included estimating diabetes and hypertension disease risk at health district, constituency and regional level in Namibia, estimating macro determinants of diabetes and hypertension and exploring various disease mapping approaches to model fitting of diabetes and hypertension in Namibia.

The subsequent subsections covers the discussion of results found using some of the disease mapping approaches used in the analysis of diabetes and hypertension. Moreover, the results from the multiscale models of diabetes and hypertension for the study are discussed.

5.1.1 Results of the disease mapping methods of the analysis

The distribution of total observed counts of diabetes and hypertension, RR estimates from three BYM models fitted for both diabetes and hypertension as well as SMR maps produced were some of the approaches used as descriptive statistics in the study.

The total observed counts of diabetes and hypertension merely provided the total number of observed counts at each region, health district and constituency in Namibia. SMR maps produced was used to reduce the noise in the data to provide more plausible maps while three BYM models fitted for each disease addressed the problem large random variability and producing unstable estimates of RR. In addition, to generate locally smoothed RR estimates of diabetes and hypertension using the CAR model.

Furthermore the maps produced from the three BYM models fitted served as the basis to measure the spatial autocorrelation or pattern in the data.

With respect to total observed counts of diabetes and hypertension, results from the study indicate that Khomas region had the highest reported cases at regional level. A similar pattern was observed from the analysis at health district level whereby Windhoek health district was found to have the highest observed counts of diabetes and hypertension. Similarly, at constituency level, the total observed counts of diabetes and hypertension was found to be highest in Khomasdal North constituency.

The results of the total observed counts of diabetes and hypertension at region, health district and constituency level are not surprising as they merely provide total reported cases of diabetes and hypertension between 2008 and 2014. However, this also suggests that a large number of the population under study reported cases of diabetes and hypertension at the Windhoek health districts between 2008 and 2014. At the regional level, results implies that more cases of diabetes and hypertension were reported in Khomas region while at constitituency level the large number of cases of diabetes and hypertension were reported in Khomasdal North constituency.

The study also shows areas where the total number of observed cases of diabetes and hypertension were low. At regional level, Ohangwena and Kavango region had the lowest number of cases of diabetes while Hardap and Karas region was also found to have the lowest number of total observed counts of hypertension. At health district level, Aranos health district had the lowest counts of diabetes implying that the health

district had the lowest number of cases reported from 2008 to 2014. For hypertension, the lowest was in Aranos and Karasburg health district.

Again, the low observed counts of diabetes and hypertension suggests that this regions were among the regions with a low number of cases reported from 2008 to 2014 indicating that there were less cases of diabetes and hypertension observed in the regions.

For total observed counts at constituency level, results indicate that overall there were not many constituencies with a high number of observed counts of diabetes and hypertension suggesting that there were not many cases of diabetes and hypertension from 2008 to 2014 in many constituencies. The highest observed counts of diabetes and hypertension were however found to be in Khomasdal North constituency suggesting that even though most of the constituencies had low cases of diabetes and hypertension reported, the constituency had the highest cases reported over the 7 year period.

The results from the study on the distribution of total observed counts of diabetes and hypertension are consistent with what has been reported in other studies. For instance, the study found the highest number of observed counts to be in Windhoek health district, Khomas region and Khomasdal North constituency which is similar with results by the MoHSS who found the incidence of diabetes to be highest in Khomas region between 2008 and 2013 (MoHSS, 2014).

With regard to hypertension, the MoHSS also found hypertension to be highest among women and men living in Khomas region (57 percent) (MoHSS & ICF International, 2014). Moreover, the MoHSS also report that hypertension has been increasing steadily over the years with Khomas region with the highest number of reported cases of hypertension between 2008 and 2013 which is consistent with the current study. Furthermore, the study also found Karas region to have the lowest reported cases of hypertension. Similarly, the MoHSS report that Karas region had the lowest cases of hypertension between 2008 and 2013 (MoHSS, 2014).

Mapping the geographical distribution of total counts of diabetes and hypertension however is not sufficient as the variability in the data is not accounted for and the data contains noise. However, the SMR method commonly used in disease mapping can assist in reducing noise in the data. Moreover, geographical dispersion of a disease can also be identified.

Results from the study using the SMR method showed significant spatial variation at the different levels. Ideally, an SMR greater than one suggests an increased risk of diabetes and hypertension in an area while an SMR less than one implies that there is a reduced risk of diabetes and hypertension in the area.

The results shows that Kunene region had the highest SMR of diabetes suggesting an increased risk of diabetes in the area. The same result was evident for hypertension, where Kunene and Ohangwena region was found to have the highest SMR indicating an increased risk for hypertension in the regions.

The study further found regions with a reduced risk for diabetes and hypertension, for instance, regions such as Hardap, Omusati and Oshana region were found to have the lowest SMR of diabetes while Karas region had the lowest SMR of hypertension.

Results from the study also give SMR estimates at each health districts, several health districts were found to have a high SMR of diabetes and hypertension. Notably, Khorixas health district and Nyangana health district was found to have the highest SMR of diabetes and hypertension respectively suggesting that there was an increased risk for diabetes and hypertension in the health districts.

Results from the study also show that some of the health districts had a reduced risk for diabetes and hypertension. For instance, health districts such as Aranos and Nankudu were among found to exhibit the lowest SMR of diabetes while Karasburg was found to have the lowest SMR.

Estimates of SMR at constituency level also showed some constituencies to exhibit a high SMR. For instance, the SMR for diabetes was evident in Khomasdal North constituency. Again, suggesting an increased risk of diabetes in the constituency. With regard to hypertension, Aminuis, Eenhana, Khorixas, Oniipa, Okongo, Rehoboth Rural and Steinhausen were among found to have a relatively high SMR suggesting that these constituencies among others had an increased risk for hypertension.

In general however, most constituencies however were found to have a relatively low SMR of diabetes and hypertension when the SMR method was used, implying that there was generally a reduced risk of diabetes and hypertension in most constituencies.

The current study found results that were consistent with what has been reported by Meza (2003), whereby most constituencies exhibited a low SMR suggesting that the areas had a reduced risk for diabetes and hypertension which might not have been the case as most of this areas with a low SMR were among areas with the highest total observed counts. Furthermore, at health district level the SMR method also proved to provide unstable estimates as suggested by Lawson et al. (2003). For example, Windhoek health district had the highest number of observed counts of diabetes and hypertension (15462 and 30620 respectively) but was among the health districts with the lowest SMR. Moreover, Khomas region was also found to have the highest observed counts of diabetes and hypertension. When the SMR method was used as a measure of RR, regions with the highest observed counts were found to exhibit a low SMR.

In general, relying on the SMR method to measure RR can badly misrepresent the geographical distribution of disease because no account is taken of varying population size over the map. Moreover, estimated SMRs especially where a few based observed counts exist may be the extremes of the map, and hence dominate its pattern. Similarly, the convolution model (BYM model) does not accommodate a spatial scaling effect associated with aggregation in the data space Aregay et al. (2015).

The results of the study using the Besag model were not surprising because they showed a spatial pattern at region, health district or constituency level. That is, the entire maps appeared to displays clustering in mostly the southern and central parts of Namibia suggesting that regions, health districts or constituencies near one another

had similar RR estimates. For instance, results indicate that the RR of diabetes and hypertension were high in regions such as Khomas, Karas, Erongo, Hardap, Omaheke and Otjozondjupa among others while regions that had a low RR were only evident in Oshana for diabetes and Oshana and Oshikoto for hypertension.

A similar pattern was observed for most health districts and at constituencies. For example, results indicate that diabetes and hypertension appeared to be high in health districts such as Karasburg, Keetmanshoop and Luderitz, Aranos, Gobabis, Rundu, Mariental, Walvis Bay, Windhoek, Rehoboth, Swakopmund and Grootfontein among others. Furthermore, at constituency level, the RR of diabetes and hypertension was high in the southern part of Namibia, specifically constituencies situated in the southern part of the region. For instance, Luderitz, Orangemund, Karasburg, Berseba, Gibeon, Keetmanshoop Rural, Marienatal Rural, Mariental Urban, Walvis Bay Rural, Rehoboth Rural, Aminuis, Gobabis, Windhoek Rural, Windhoek West, Windhoek East, Khomasdal North, Samora Machel, Tobias Hainyeko, Soweto, Moses Garoeb, Katutura East, Katutura Central and Rehoboth East Urban appeared to show a relatively high RR among others.

As expected, the method proposed by Besag at al. (1991) to study spatial patterns has its drawbacks because the method does not accommodate a spatial scaling effect associated with aggregation in the data space as stated earlier by Aregay et al. (2015). Hence, the study is consistent with the statement by Aregay et al. (2015) on using the BYM model to study the spatial variation in disease because the current study found

results where there was a positive autocorrelation mainly in the southern and central parts of Namibia.

One major advantage of using a multiscale model concerns the scaling effect which is viewed important in disease mapping, Aregay et al. (2015) argues that the scale effect reduces the variability of the data and makes them more alike, and hence the results from the finer data may not be consistent with the results from the coarser data.

The multiscale approach was discussed earlier in chapter 2, the approach mainly involves factorizing the likelihood into the individual components of local information (Kolaczyk & Haung, 2001). The subsequent section discusses the results of the study from the multiscale models of diabetes and hypertension that were fitted in the study.

5.1.2 Fixed effects of diabetes

The covariates were added to the multiscale model of diabetes and hypertension selected based on the ΔDIC. In the analysis the multiscale model with covariates was used for diabetes, because adding the covariates improved the model. The multiscale model for hypertension did not improve when covariates were added to the model because the ΔDIC increased. Furthermore, the ΔDIC was also found to be greater than four ($\Delta DIC \geq 4$ implies that the models are significantly different). The study did not interpret any effects of covariates on hypertension since the best fit model selected had no covariates.

Although adding covariates to M_{11} improved the model for diabetes, results from the study indicate that the effect of safe water (RR= 1.0046, 95 % CI: 0.9905, 1.0191), the

effect of wood/charcoal (RR= 1.0051, 95% CI: 0.9945, 1.0160), the effect of wages and salaries as main source of income (RR= 1.0104, 95% CI: 0.9961, 1.0249) and the effect of pension as the main source of income (RR= 1.0007, 95 % CI: 0.9812, 1.0219) was not statistically associated with diabetes.

In addition, the effect of education attainment: incomplete primary education (RR= 0.9901, 95% CI: 0.9653, 1.0129) was also not found to be statistically associated with diabetes which might require further investigation. It is well known that higher levels of educational attainment are associated with better employment and increased earning potential, which provides greater access to health care (Ross & Wu, 1995). In addition, those with higher levels of educational attainment are more likely to engage in healthier behaviors, receive preventative care, and less likely to have adverse outcomes that result in death (Hummer & Lariscy, 2011). Therefore, individuals with lower educational attainment are more likely to have heart disease, diabetes, asthma, and ultimately die earlier compared with those with higher educational attainment levels (Sacco, Bykowski, Mayhew & White, 2012; Whitaker et al., 2014).

Hence, all covariates were not statistically associated with diabetes and therefore did not have any significant impact on determining diabetes. In general, one would want the RR of an effect to be less than one and the corresponding lower and upper credible interval to be less than one, which would imply that the there is a decreased risk of diabetes. However, an RR greater than one with the corresponding lower and upper credible intervals greater than one would also imply that there is an increased

association between the effect and diabetes and that the effect is statistically significant.

5.1.3 Spatial effects of diabetes and hypertension

Spatial effects for diabetes were presented as RR at different levels considered in the study. An RR greater than one implies that there is an increased risk for diabetes and hypertension in the area while an RR less than one implies that there is a decreased risk of diabetes and hypertension in an area. Furthermore, the RR maps of diabetes and hypertension given for the analysis at different levels using the multiscale models fitted suggests that spatial variation exists among the different areas.

The study modelled diabetes and hypertension one at a time, using the DIC for model comparison. Ideally, the model with the lowest DIC is considered the best fit model. The study however applied the rule of thumb suggested by Best and Richardson (2009, 2012) on model selection (Neema & Böhning, 2012). As a result, it was reasonable to only fit M_{12} and M_{23} in order to fully evaluate the effects of health district, region and constituency of the structured and unstructured heterogeneity on the RR of the two diseases.

In terms of spatial variation amongst the regions in Namibia, Kavango region had the highest RR (RR= 1.027) of diabetes while Khomas region (RR= 1.019) had the highest RR of hypertension suggesting that there was an increased risk for diabetes and hypertension respectively in those region.

Regions such as Caprivi, Khomas and Otjozondjupa were also found to have an increased risk for diabetes. Although results also shows that regions such as Omusati and Oshikoto appeared to have a low RR of diabetes, suggesting that there was a decreased risk for diabetes in those areas. The lowest RR was however found in Oshana region (RR=0.969) suggesting that a reduced risk for diabetes was also evident in the region. The study found consistencies with what has been found by the MoHSS and ICF (2014) where for instance, Khomas and Otjozondjupa region was among the regions with a high percentage of men and women informed by a health professional that they had diabetes while Oshana region was reported to be among the regions with a low percentage of men and women informed that they had diabetes which is consistent with the current study.

Results from the study also indicate that Erongo, Hardap, Otjozondjupa, Kavango and Karas region were among the regions with a high RR of hypertension, suggesting an increased risk of hypertension in those regions. However, the lowest RR of hypertension according to the results of the study was found to be in Oshikoto (RR= 0.976). In addition, Oshana region was also found to have a low RR of hypertension suggesting that there was a reduced risk for hypertension in the two regions. The results for hypertension are not surprising since the MoHSS and ICF international (2014) also found Khomas, Erongo, Hardap and Kavango region among others to constitute regions where most men and women have been told by a health professional that they had hypertension which appears to be consistent with the current study. In terms of reduced risk for hypertension, there was some consistencies found with

Oshana region that also appeared to have one of the lowest percentages of men and women being informed by a health professional that they had hypertension (MoHSS & ICF international, 2014).

At health district level, Rundu health district (RR= 1.037) was found to have the highest RR of diabetes suggesting that there was an increased risk of diabetes in the health district. Other health districts with an increased risk of diabetes were Andara, Grootfontein and Okakarara among others. Results further indicate that Oshakati health district (RR= 0.967) had the lowest RR of diabetes. Engela, Eenhana and Oshikuku were also found to have a low RR of diabetes suggesting that there is a reduced risk of diabetes in the health districts. For hypertension, Andara (RR=1.012) had the highest RR. Other health districts with a high RR were Rundu, Swakopmund and Windhoek suggesting that an increased risk for hypertension were in these health districts. However, Oshakati health district (RR= 0.975) was found to have the lowest RR indicating that a reduced risk for hypertension was in the health district.

Results further indicate that at constituency level, an increased risk for diabetes were mainly in Rundu Urban constituency (RR= 1.064) which was found to have the highest RR. Other constituencies that appeared to exhibit a high RR were Mashare, Kapako, Katima Mulilo Urban, Keetmanshoop Urban, Khomasdal North, Khorixas, Katutura Central and Okakarara among others. The results further indicate that Rehoboth West Urban was among the constituencies with the lowest RR of diabetes (RR= 0.972) indicating that a reduced risk for diabetes was in the constituency. Other constituencies such as Oranjemund, Keetmashoop Rural, Uuvudhiya, Eengondi and Onayena among

others were also found to have a low RR of diabetes. Again, suggesting a reduced or decreased risk for diabetes.

With respect to hypertension at constituency level, an increased risk for hypertension were in constituencies such as Rundu Urban, Katima Mulilo Urban, Khomasdal North, Walvis Bay Urban, Swakopmund, Rehoboth East Urban, Katutura East, Keetmanshoop Urban, Mariental Urban and Tobias Hainyeko among others with Rundu Urban (RR= 1.03) found to have the highest RR of hypertension compared to the other constituencies mentioned. Results further indicate that there were also constituencies with low RR of hypertension, among others were Ompundja, Okatyali, Eengondi, Omuthiya Qwiimpundi, Olukonda, Ohangwena, Etayi, Oshakati East, Endola, Okaku and Kalahari with the lowest found in Ompundja (RR= 0.967). The low RR of hypertension in these constituencies suggests that there is a decreased risk for hypertension in the constituencies.

5.2 Limitations of the study

The study had several limitations. Firstly, the data source only covered cases that were reported in the health facilities in Namibia implying that cases that were not reported at the health facilities were not part of the study. Therefore those who chose self-medication or used other means apart from public hospitals were not covered by the study. Secondly, the Namibia HIS did not collect information on behavior and habit factors such as smoking, alcohol consumption, health lifestyle including physical activity and eating habits which may be considered important in measuring the impacts of socio-economic factors on diabetes and hypertension. Thirdly, literature suggests

using recent self-reported data, for instance DHS data employed in some studies such as those of Kazembe and Mpeketula (2009), Kazembe and Kandala (2015), Manda (2011), Kandala et al. (2013). Hence, the study may not be representative to the situation on the ground. Lastly, there were no geo-coordinates available at health facility level, therefore the analysis could not be done at the level.

5.3 Conclusion

In terms of multiscale modelling of diseases risk in Namibia, the availability of limited literature for the study suggests that this is the first study that attempted to estimate the geographical variation of diabetes and hypertension at the region, health district and constituency level in Namibia.

As expected, the spatial analysis section shows that there is high risk across the different levels considered in the study. Furthermore, the multiscale model for diabetes and hypertension applied showed no clusters or spatial pattern which was not the case when the Besag model was used in the analysis. However, the areas found to have elevated or increased risk for diabetes and hypertension might need to be prioritized. Furthermore, areas found to have a reduced risk for diabetes and hypertension might requires future control interventions related to monitoring and evaluation.

It is therefore hoped that the study will assist policy makers particularly those involved in health planning to develop comprehensive programmes or targeted interventions particularly in areas that were found to have elevated risk of diabetes and hypertension, in turn developing programmes and strategies aiming at improving the health and well-

being of the population. Moreover, the study hopes that considering spatial factors in planning of health programmes related to diabetes and hypertension could assist in the achievement of National development Plans (NDPs) goals such as those outlined in the NDP 4 in line with the progress towards the international goals such as UN Millennium Development Goals.

Because of the complications of the data used for the study which was affected by incompleteness, the study could in the future be replicated or a similar study could be conducted using population based data obtained from the recent 2013 NDHS survey undertaken in Namibia, in addition measurement error models could also be considered to improve on the current study. Moreover, the possibility of combining different survey datasets to generate better maps of RR could be explored. Furthermore, since recent studies have also been focusing on modelling disease outcomes jointly (Manda, 2011), the study recommends to consider jointly modelling diabetes and hypertension whereby instead of modelling both diseases separately, both disease are modelled at the same time.

REFERENCES

Aregay, M., Lawson, A. B., Faes, C., & Kirby, R. (2014). Bayesian multiscale modeling for aggregated disease mapping data. *In Proceedings of the Third ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health* (pp. 45-48). ACM.

Belitz, C., Brezger, A., Kneib, T., & Lang, S. (2009). *BayesX-software for Bayesian Inference in Structure Additive Regression Models,* version 2.01. Retrieved October 10, from http://www.dtst.uni-muechen.de/bayes2.01.

Belue, R., Okoror, T. A., Iwelunmor, J., Taylor, K. D., Degboe, A. N., Agyemang, C., & Ogedegbe, G. (2009). An overview of cardiovascular risk factor burden in Sub-Saharan African countries: a socio-cultural perspective. *Globalization and Health*, *5*(1), 10.

Besag, J., York, J.C., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics 43*, 1-59

Best, N., & Richardson, S. (2009). Introduction to Bayesian analysis using WinBUGS. Introduction to Bayesian Analysis and WinBUGS course, Imperial College.

Best, N., & Richardson, S. (2012). Introduction to Bayesian analysis using WinBUGS. In *Workshop-Hilton Glasgow Grosvenor* (Vol. 3, pp. 08-00).

Best, N., Richardson, S., & Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical methods in medical research,14*(1), 35-59.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H. W., & Silkey, M. (1997). Natural variability of benthic species composition in the Delaware Bay. *Environmental and Ecological Statistics, 4*(2), 95-115.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9-25.

Carlin, B. P., & Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. *Bayesian Statistics, 7*, 45-63.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardised relative risks for use in disease mapping. *Biometics, 43,*671-681.

Cressie, N. (1992). Smoothing regional maps using empirical Bayes predictors. *Geographical Analysis*, *24*(1), 75-95.

Cressie, N., & Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association, 84*(406), 393-401.

De Oliveira, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, *64*(1), 107-133.

Denison, D. G. T., & Holmes, C. C. (2001). Bayesian partitioning for estimating disease risk. *Biometrics*, *57*(1), 143-149.

Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association, 70*(350), 311-319.

Gangnon, R. E., & Clayton, M. K. (2000). Bayesian detection and modeling of spatial disease clustering. *Biometrics, 56*(3), 922-935.

Gangnon, R. E., & Clayton, M. K. (2003). A hierarchical model for spatially clustered disease rates. *Statistics in Medicine, 22*(20), 3213-3228.

Gómez-Rubio, V., & López-Quılez, A. (2006). Empirical and Full Bayes estimators for disease mapping. *In International workshop on Spatio-temporal Modelling (METMA3)*. Pamplona, Spain.

Held, L., Natário, I., Fenton, S. E., Rue, H., & Becker, N. (2005). Towards joint disease mapping. *Statistical Methods in Medical Research*, *14*(1), 61-82.

Hummer, R. A., & Lariscy, J. T. (2011). Educational attainment and adult mortality. In *International handbook of adult mortality*, 241-261, Springer Netherlands.

Islam, S. M. S., Purnat, T. D., Phuong, N. T. A., Mwingira, U., Schacht, K., & Fröschl, G. (2014). Non-Communicable Diseases (NCDs) in developing countries: A symposium report. *Globalization and Health*, *10*(1), 1-8.

Jin, X., Carlin, B. P., & Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, *61*(4), 950-961.

Kandala, N. B., Tigbe, W., Manda, S. O., & Stranges, S. (2013). Geographic variation of hypertension in Sub-Saharan Africa: A case study of South Africa. *American Journal of Hypertension*, *26*(3), 382-391.

Kazembe, L. N., & Kandala, N. B. (2015). Estimating areas of common risk in low birth weight and infant mortality in Namibia: A joint spatial analysis at sub-regional level. *Spatial and Spatio-temporal epidemiology*, *12*, 27-37.

Kazembe, L. N., & Mpeketula, P. M. (2009). Detecting geographical variability in risk of malaria-attributable morbidity using spatial models. *Analysis*, *4*(5), 6.

Kim, H., Sun, D., & Tsutakawa, R. K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *Journal of the American Statistical Association*, *96*(456), 1506-1521.

Knorr-Held, L., & Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(1), 73-85.

Knorr-Held, L., & Rainer, E. (2001). Projections of lung cancer mortality in West Germany: A case study in Bayesian prediction. *Biostatistics*, *2(*1), 109-129.

Kolaczyk, E. D., & Huang, H. (2001). Multiscale statistical models for hierarchical spatial aggregation. *Geographical Analysis*, *33*(2), 95-118.

Lawson, A. (2008). Bayesian Disease Mapping: Hierarchical modelling in spatial epidemiology: CRC Press, Boca Raton, FL.

Lawson, A. B. and Clark, A. (1999). Markov chain Monte Carlo methods for clustering in case event and count data in spatial epidemiology. In M. E. Halloran and D. Berry (Eds.), *Statistics and Epidemiology: Environment and Clinical Trials*, pp. 193–218. New York: Springer Verlag.

Lawson, A. B., Biggeri, A. B., Boehning, D., Lesaffre, E., Viel, J. F., Clark, A., ... & Divino, F. (2000). Disease mapping models: An empirical evaluation. Disease Mapping Collaborative Group. *Statistics in Medicine*, *19*(17-18), 2217-2241.

Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. (2003). *Disease mapping with WinBUGS and MLwiN*, 11, John Wiley & Sons.

Lee, D., & Mitchell, R. (2013). Locally adaptive spatial smoothing using conditional auto-regressive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(4), 593-608.

Lee, K. S., & Kim, S. U. (2008). Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method. *Hydrological Processes*, *22*(12), 1949-1964.

Louie, M. M., & Kolaczyk, E. D. (2006). A multiscale method for disease mapping in spatial epidemiology. *Statistics in Medicine*, *25*(8), 1287-1306.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A Practical Introduction to Bayesian Analysis*. CRC press. Boca Raton, FL, USA.

Manda, S. (2011). Joint Mapping Modelling for Multiple Health Problems in South Africa. Retrieved from http://sacemaquarterly.com/tag/joint-mapping-modelling.

Mardia, K. V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, *24*(2), 265-284.

Marquez, P. V., & Farrington, J. L. (2013). The challenge of non-communicable diseases and road traffic injuries in sub-Saharan Africa: An overview. Washington, DC: The World Bank.

Martino, S., & Rue, H. (2010). *Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the INLA program*. Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. Compiled on April, 8, 2010.

Meza, J. L. (2003). Empirical Bayes estimation smoothing of relative risks in disease mapping. *Journal of Statistical Planning and Inference*, *112*(1), 43-62.

Micha, R., Khatibzadeh, S., Shi, P., Fahimi, S., Lim, S., Andrews, K. G., ... & Mozaffarian, D. (2014). Global Burden of Diseases Nutrition and Chronic Diseases Expert Group NutriCoDE. Global, regional, and national consumption levels of dietary fats and oils in 1990 and 2010: A systematic analysis including 266 country--specific nutrition surveys. *BMJ*, *348*, g2272.

Namibia Statistics Agency. (2015). *Namibia Labour Force Survey 2014 Report*. Namibia Statistics Agency, Windhoek.

Neema, I., & Böhning, D. (2012). Monitoring murder crime in Namibia using Bayesian space-time models. *Journal of Probability and Statistics*, *2012*.

R Core Team (2015). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Rao, J. N. (2003). Small area estimation. John Wiley & Sons, Inc. Hoboken, New Jersey.

Ross, C. E., & Wu, C. L. (1995). The links between education and health. *American sociological review*, 719-745.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319-392.

Sacco, W. P., Bykowski, C. A., Mayhew, L. L., & White, K. E. (2012). Educational attainment moderates the effect of a brief diabetes self-care intervention. *Diabetes Research and Clinical Practice*, *95*(1), 62-67.

Sain, S. R., & Cressie, N. (2002). Multivariate lattice models for spatial environmental data. In *Proc. of the ASA Section on Statistics and the Environment*, 2820-2825.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

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: Series B (Statistical Methodology)*, *64*(4), 583-639.

The Ministry of Health and Social Services (MoHSS) & Macro International Inc (2008). *Namibia Demographic and Health Survey 2006-07*. Windhoek, Namibia and Calverton, Maryland, USA: MoHSS and Macro International Inc.

The Namibia Ministry of Health and Social Services (MoHSS) and ICF International. 2014. *The Namibia Demographic and Health Survey 2013*. Windhoek, Namibia, and Rockville, Maryland, USA: MoHSS and ICF International.

The Namibia Ministry of Health and Social Services. (2014). Health Information Report 2008-2013. Windhoek, Namibia.

Torabi, M. (2014). Spatial generalized linear mixed models with multivariate CAR models for areal data. *Spatial Statistics, 10*, 12-26.

Torabi, M., & Rosychuk, R. J. (2012). Hierarchical Bayesian spatiotemporal analysis of childhood cancer trends. *Geographical Analysis*, *44*(2), 109-120.

Tu, Y. K., & Greenwood, D. C. (Eds.). (2012). *Modern methods for epidemiology*. Springer Science & Business Media.

Tzala, E., & Best, N. (2007). Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. *Statistical Methods in Medical Research*. *17*(1):97– 118.

Ugarte, M. D., Goicoa, T., & Militino, A. F. (2009). Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics & Data Analysis, 53*(8), 2938-2949.

Ugarte, M. D., Militino, A. F., & Goicoa, T. (2008). Prediction error estimators in empirical Bayes disease mapping. *Environmetrics*, *19*(3), 287-300.

Waller, L. A., & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*, 368, John Wiley & Sons.

Waller, L.A., Carlin, B.P., Xia, H., Gelfand, A.E., 1997. Hierarchical spatio-temporal mapping of disease rates. J. Amer. Statist. Assoc. 92, 607–617.

Whitaker, S. M., Bowie, J. V., McCleary, R., Gaskin, D. J., LaVeist, T. A., & Thorpe, R. J. (2014). The association between educational attainment and diabetes among men in the United States. *American journal of men's health*,*8*(4), 349-356.

World Health Organisation-Namibia (2010). *Non-Communicable Diseases (NCDs) Namibia's Silent Killers*. Windhoek, World Health Organisation-Namibia.

World Health Organization. (2014). *Global Status Report on Noncommunicable Diseases 2013*. Geneva: World Health Organization.

Zhu, L., & Carlin, B. P. (2000). Comparing hierarchical models for spatio-temporally misaligned data using the deviance information criterion. *Statistics in Medicine*, *19*(17-18), 2265-2278.

ANNEXURE A: BayesX codes

BayesX is the software for Bayesian Inference in Structured Additives Regression Models and is the software used for spatial analysis for this thesis. BayesX is free software that can be downloaded from the site (http://www.stat.uni-muenchen.de/~bayesx/). Version 2.1 of BayesX developed in (07.05.2012) was used for this thesis. Version 2.1 of BayesX permits Bayesian Inference on Markov Chain Monte Carlo simulation techniques. For all the models fitted in BayesX version 2.1, 52000 iterations were run with a burn of 2000 for each model.

Below is the syntax used for fully Bayesian for the study:

Below is the syntax used for fully Bayesian for the study:

```
> dataset multiscale
> multiscale.infile using
C:\Users\Tharris.NSA\Desktop\School_may\multiscale_data_final.txt
> multiscale.describe
> map hdmap
> hdmap.infile,graph using
C:\Users\Tharris.NSA\Desktop\School_may\neibours_changes1.txt
> map constimap
> constimap.infile using C:\Users\Tharris.NSA\Desktop\School_may\namibia.csv
```

```
> map regmap
> regmap.infile using C:\Users\Tharris.NSA\Desktop\School_may\nam_regions.csv
```

(1) Codes used to fit models of diabetes at health district, region and constituency level

```
> bayesreg m1
> m1.regress log_diab=safe_water+wood_coal+wages+pension+primary,
family=poisson predict using multiscale
> bayesreg m2
> m2.regress log_diab=hd(spatial, map=hdmap), family=poisson predict using
multiscale
> bayesreg m3
> m3.regress log_diab=const(spatial, map=constimap), family=poisson predict using
multiscale
> bayesreg m4
> m4.regress log_diab=reg(spatial, map=regmap), family=poisson predict using
multiscale
> bayesreg m5
> m5.regress log_diab=hd(spatial, map=hdmap)+hd(random), family=poisson
predict using multiscale
> bayesreg m6
> m6.regress log_diab=const(spatial, map=constimap)+const(random),
family=poisson predict using multiscale
```

> bayesreg m7

> m7.regress log_diab=reg(spatial, map=regmap)+reg(random), family=poisson predict using multiscale

> bayesreg m8

> m8.regress log_diab=hd(spatial, map=hdmap)+const(spatial,map=constimap), family=poisson predict using multiscale

> bayesreg m9

> m9.regress log_diab=hd(spatial, map=hdmap)+reg(spatial,map=regmap), family=poisson predict using multiscale

> bayesreg m10

> m10.regress log_diab=hd(spatial,map=hdmap)+hd(random)+const(spatial,map=constimap), family=poisson predict using multiscale

> bayesreg m11

> m11.regress log_diab=const(spatial,map=constimap)+const(random)+hd(spatial,map=hdmap)+reg(spatial,map=regmap), family=poisson predict using multiscale

> bayesreg m23

> m23.regress log_diab=const(spatial,map=constimap)+const(random)+hd(spatial,map=hdmap)+reg(spatial, map=regmap)+safe_water+wood_coal+wages+pension+primary, family=poisson predict using multiscale

(2) Codes used to fit models of hypertension at health district, region and constituency level

```
> bayesreg m12
> m12.regress log_ht=safe_water+wood_coal+wages+pension+primary,
family=poisson predict using multiscale
> bayesreg m13
> m13.regress log_ht=hd(spatial, map=hdmap), family=poisson predict using
multiscale
> bayesreg m14
> m14.regress log_ht=const(spatial, map=constimap), family=poisson predict using
multiscale
> bayesreg m15
> m15.regress log_ht=reg(spatial, map=regmap), family=poisson predict using
multiscale
> bayesreg m16
> m16.regress log_ht=hd(spatial, map=hdmap)+hd(random), family=poisson predict
using multiscale
> bayesreg m17
> m17.regress log_ht=const(spatial, map=constimap)+const(random),
family=poisson predict using multiscale
> bayesreg m18
> m18.regress log_ht=reg(spatial, map=regmap)+reg(random), family=poisson
```

predict using multiscale

> bayesreg m19

> m19.regress log_ht=hd(spatial, map=hdmap)+const(spatial,map=constimap), family=poisson predict using multiscale

> bayesreg m20

> m20.regress log_ht=hd(spatial, map=hdmap)+reg(spatial,map=regmap), family=poisson predict using multiscale

> bayesreg m21

> m21.regress log_ht=hd(spatial,map=hdmap)+hd(random)+const(spatial,map=constimap), family=poisson predict using multiscale

> bayesreg m22

> m22.regress log_ht=const(spatial,map=constimap)+const(random)+hd(spatial,map=hdmap)+reg(spatial,map=regmap), family=poisson predict using multiscale

> bayesreg m24

> m24.regress log_ht=const(spatial,map=constimap)+const(random)+hd(spatial,map=hdmap)+reg(spatial, map=regmap)+safe_water+wood_coal+wages+pension+primary, family=poisson predict using multiscale

(3) Codes used to map the observed counts of diabetes and hypertension at a health district, regional and constituency level

(a) Codes for observed counts of diabetes and hypertension at health district and regional level:

```
> dataset hd
> hd.infile using C:\Users\Tharris.NSA\Documents\NCD_final\24included.txt
> map hdbound
> hdbound.infile using C:\Users\Tharris.NSA\Documents\NCD_final\hdbound.csv
> map regmap
> regmap.infile using
C:\Users\Tharris.NSA\Documents\NCD_final\nam_regions.csv
> graph diab_observed_hd
> diab_observed_hd.drawmap diab hd, map=hdbound swapcolors using hd
> graph ht_observed_hd
> ht_observed_hd.drawmap ht hd, map=hdbound swapcolors using hd
> graph diab_region_observed
> diab_region_observed.drawmap diab reg, map=regmap swapcolors using hd
> graph ht_region_observed
> ht_region_observed.drawmap ht reg, map=regmap swapcolors using hd
```

(b) Codes used to map observed counts of diabetes and hypertension at constituency:

```
> dataset consti
> consti.infile using
C:\Users\Tharris.NSA\Documents\NCD_final\consti_final_january_log_rounded.tx
> map constimap
```

> constimap.infile using C:\Users\Tharris.NSA\Documents\NCD_final\namibia.csv

> graph diab_consti_dist

> diab_consti_dist.drawmap diab const,map=constimap swapcolors using consti

> graph ht_consti_dist

> ht_consti_dist.drawmap ht const,map=constimap swapcolors using consti

(4) Codes used to map the SMR of diabetes and hypertension at a health district, regional and constituency level

(a) Codes for SMR of diabetes and hypertension at health district and regional level:

> graph smr_diab_hd

> smr_diab_hd.drawmap smr_diab hd, map=hdbound swapcolors using hd

> graph smr_ht_hd

> smr_ht_hd.drawmap smr_ht hd, map=hdbound swapcolors using hd

> graph smr_diab_reg

> smr_diab_reg.drawmap smr_diab reg, map=regmap swapcolors using hd

> graph smr_ht_reg

> smr_ht_reg.drawmap smr_ht reg, map=regmap swapcolors using hd

(b) Codes for SMR of diabetes and hypertension at constituency level:

> dataset consti_smr

> consti_smr.infile using

C:\Users\Tharris.NSA\Documents\NCD_final\consti_final_january_log_rounded.txt

> graph diab_smr_consti1

> diab_smr_consti1.drawmap smr_diab_rounded const,map=constimap swapcolors

using consti_smr

> graph ht_smr_consti

> ht_smr_consti.drawmap smr_ht_rounded consti,map=constimap swapcolors using consti_smr

(5) Codes for structured and unstructured spatial effects of diabetes at health district, region and constituency level:

> map hdmap1

> hdmap1.infile using C:\Users\Tharris.NSA\Desktop\School_may\hdbound.csv

> map constimap

> constimap.infile using C:\Users\Tharris.NSA\Desktop\School_may\namibia.csv

> map regmap

> regmap.infile using C:\Users\Tharris.NSA\Desktop\School_may\nam_regions.csv

> dataset RR_diab

> RR_diab.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\hd_spatial_m23.txt

> graph hd_spatial

> hd_spatial.drawmap RR hd, map=hdmap1 swapcolors using RR_diab

> dataset RR_diab1

> RR_diab1.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\const_spatialtotal_m23.txt

> graph const_spatial

> const_spatial.drawmap RR const, map=constimap swapcolors using RR_diab1

> dataset RR_diab2

> RR_diab2.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\reg_spatial_m23.txt

> graph reg_spatial

> reg_spatial.drawmap RR reg, map=regmap swapcolors using RR_diab2

> dataset RR_diab3

> RR_diab3.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\const_random_m23.txt

> graph const_random

> const_random.drawmap RR const, map=constimap swapcolors using RR_diab3

(5) Codes for unstructured and structured spatial effects of hypertension at health district, region and constituency level:

> map hdmap1

> hdmap1.infile using C:\Users\Tharris.NSA\Desktop\School_may\hdbound.csv

> map constimap

> constimap.infile using C:\Users\Tharris.NSA\Desktop\School_may\namibia.csv

> map regmap

> regmap.infile using C:\Users\Tharris.NSA\Desktop\School_may\nam_regions.csv

> dataset RR_ht

> RR_ht.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\const_random_m22.txt

> graph const_random_f

> const_random_f.drawmap RR const, map=constimap swapcolors using RR_ht

> dataset RR_ht1

> RR_ht1.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\const_spatialtotal_m22.txt

> graph const_spatial_m22

> const_spatial_m22.drawmap RR const, map=constimap swapcolors using RR_ht1

> dataset RR_ht2

> RR_ht2.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\hd_spatial_m22.txt

> graph hd_spatial_m22

> hd_spatial_m22.drawmap RR hd, map=hdmap1 swapcolors using RR_ht2

> dataset RR_ht3

> RR_ht3.infile using

C:\Users\Tharris.NSA\Desktop\School_may\new_final\reg_spatial_m22.txt

> graph reg_spatial_m22

> reg_spatial_m22.drawmap RR reg, map=regmap swapcolors using RR_ht3

(6) Codes used to fit BYM models of diabetes and hypertension at a health district, region and constituency level:

> dataset car_st

> car_st.infile using C:\Users\Tharris.NSA\Documents\NCD_final\hd_final_smr.txt

> map hdmap

> hdmap.infile,graph using

C:\Users\Tharris.NSA\Documents\NCD_final\neibours_changes.txt

> bayesreg diab_car_hd

> diab_car_hd.regress log_diab=hd(spatial,map=hdmap),family=poisson predict using car_st

> bayesreg diab_car_reg

> map regmap

> regmap.infile using C:\Users\Tharris.NSA\Documents\NCD_final\nam_regions.csv

> diab_car_reg.regress log_diab=reg(spatial,map=regmap),family=poisson predict using car_st

> bayesreg ht_car_hd

> ht_car_hd.regress log_ht=hd(spatial,map=hdmap),family=poisson predict using car_st

> bayesreg ht_car_reg

> ht_car_reg.regress log_ht=reg(spatial,map=regmap),family=poisson predict using car_st

> dataset const_car_st

> const_car_st.infile using C:\Users\Tharris.NSA\Documents\NCD_final\consti_final_data1.txt

> map constimap

> constimap.infile using C:\Users\Tharris.NSA\Documents\NCD_final\namibia.csv

> bayesreg diab_car_const

> diab_car_const.regress log_diab=const(spatial,map=constimap),family=poisson

predict using const_car_st

> bayesreg ht_car_const

> ht_car_const.regress log_ht=const(spatial,map=constimap),family=poisson predict using const_car_st

(a) Codes used to map the spatially structured effects of diabetes and hypertension at a health district, region and constituency level:

> map hdmap1

> hdmap1.infile using C:\Users\Tharris.NSA\Documents\NCD_final\hdbound.csv

> dataset diab_1

> diab_1.infile using C:\Users\Tharris.NSA\Documents\NCD_final\Tommy FINAL_ESTIMATES_31_JANUARY\diab_hd_only.txt

> graph diab_hd_map

> diab_hd_map.drawmap RR hd,map=hdmap1 swapcolors using diab_1

> dataset diab_2

> diab_2.infile using C:\Users\Tharris.NSA\Documents\NCD_final\Tommy FINAL_ESTIMATES_31_JANUARY\diab_reg_only.txt

> graph diab_reg_map

> diab_reg_map.drawmap RR reg,map=regmap swapcolors using diab_2

> dataset ht_1

> ht_1.infile using C:\Users\Tharris.NSA\Documents\NCD_final\Tommy FINAL_ESTIMATES_31_JANUARY\ht_hd_only.txt

> graph ht_hd_map

> ht_hd_map.drawmap RR hd,map=hdmap1 swapcolors using ht_1

> dataset ht_2

> ht_2.infile using C:\Users\Tharris.NSA\Documents\NCD_final\Tommy FINAL_ESTIMATES_31_JANUARY\ht_reg_only.txt

> graph ht_reg_map

> ht_reg_map.drawmap RR reg,map=regmap swapcolors using ht_2

> dataset diab_3

> diab_3.infile using C:\Users\Tharris.NSA\Documents\NCD_final\Tommy FINAL_ESTIMATES_31_JANUARY\diab_const_only.txt

> graph diab_const_map

> diab_const_map.drawmap RR const,map=constimap swapcolors using diab_3

> dataset diab_31

> diab_31.infile using C:\Users\Tharris.NSA\Documents\NCD_final\Tommy FINAL_ESTIMATES_31_JANUARY\diab_const_only1.txt

> graph diab_const_map1

> diab_const_map1.drawmap RR const,map=constimap swapcolors using diab_31

> dataset ht_3

> ht_3.infile using C:\Users\Tharris.NSA\Documents\NCD_final\Tommy FINAL_ESTIMATES_31_JANUARY\ht_const_only.txt

> graph ht_const_map

> ht_const_map.drawmap RR const,map=constimap swapcolors using ht_3

ANNEXURE B: R codes

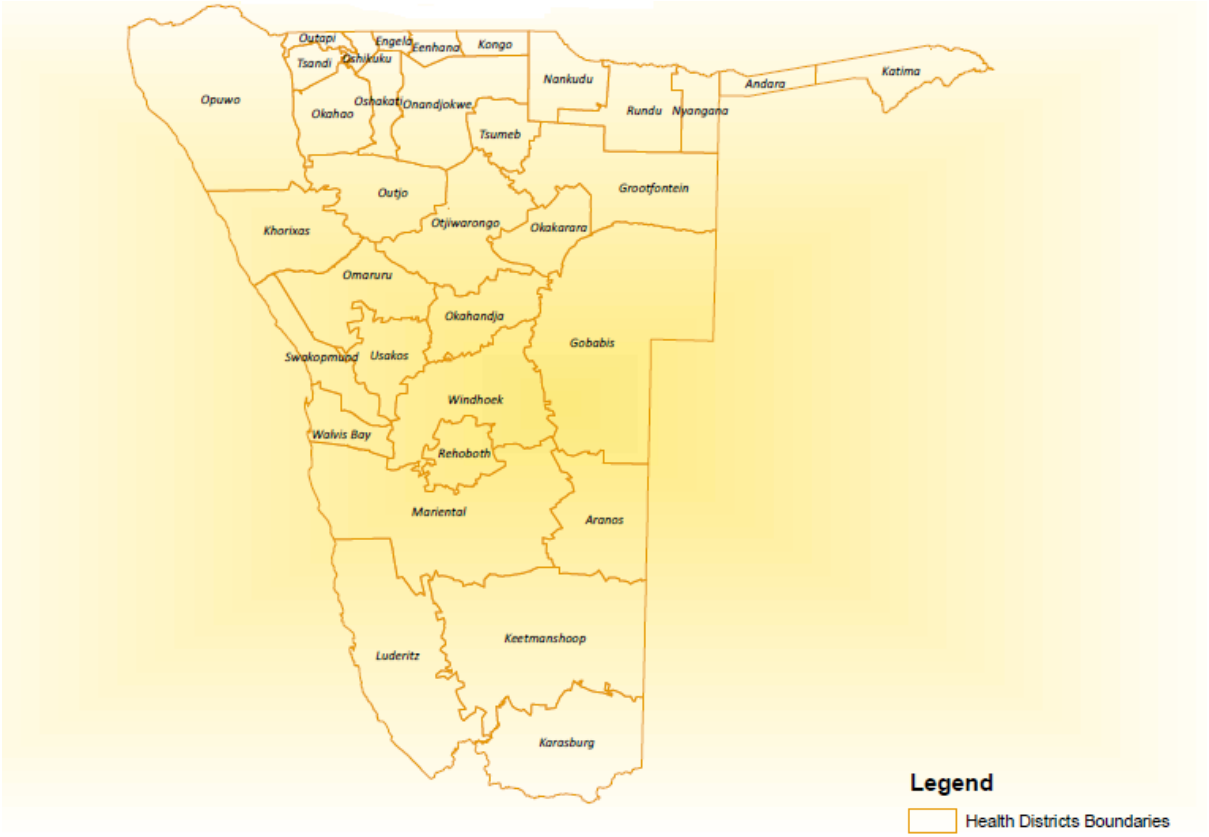
This is a freely available statistical package and is the software used for the explanatory analysis of this thesis. The R software can be freely downloaded from the site (http://www.r-project.org/.)

Below is the R syntax used for the thesis:

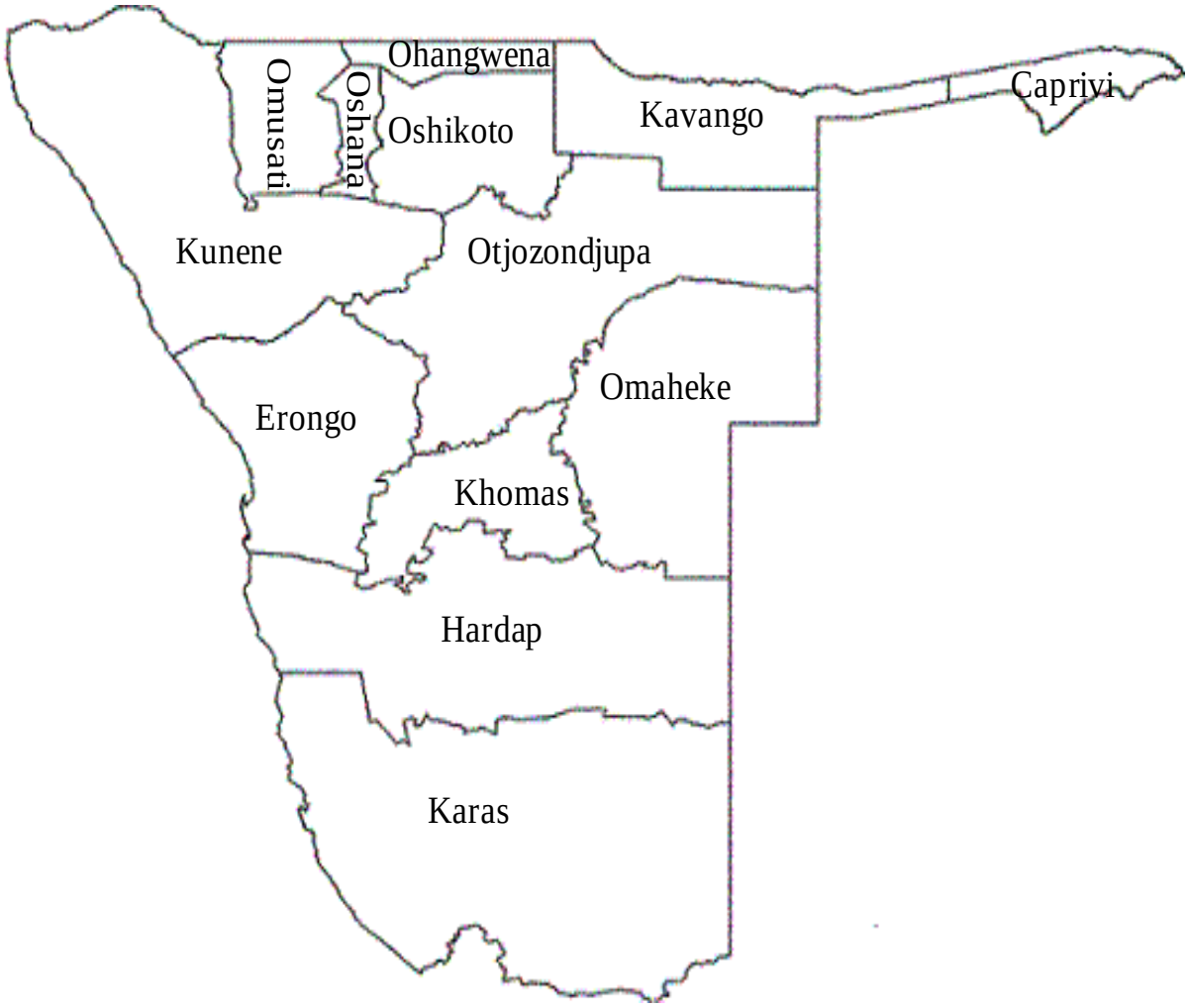
Codes used for residual boxplots for model 1-model 24

```
> diab
<read.table("C:\\Users\\Tharris.NSA\\Desktop\\School_may\\estimates\\residuals_m
1_m12.csv",header=TRUE,sep=",")
> attach(diab)
> names(diab)
> boxplot(residual~model,data=diab,xlab="Fitted models")
> abline(h=0)
> detach(diab)
> ht<
read.table("C:\\Users\\Tharris.NSA\\Desktop\\School_may\\estimates\\residuals_m12
_m24.csv",header=TRUE,sep=",")
> attach(ht)
> names(ht)
> boxplot(residual~model,data=ht,xlab="Fitted models")
> abline(h=0)
```

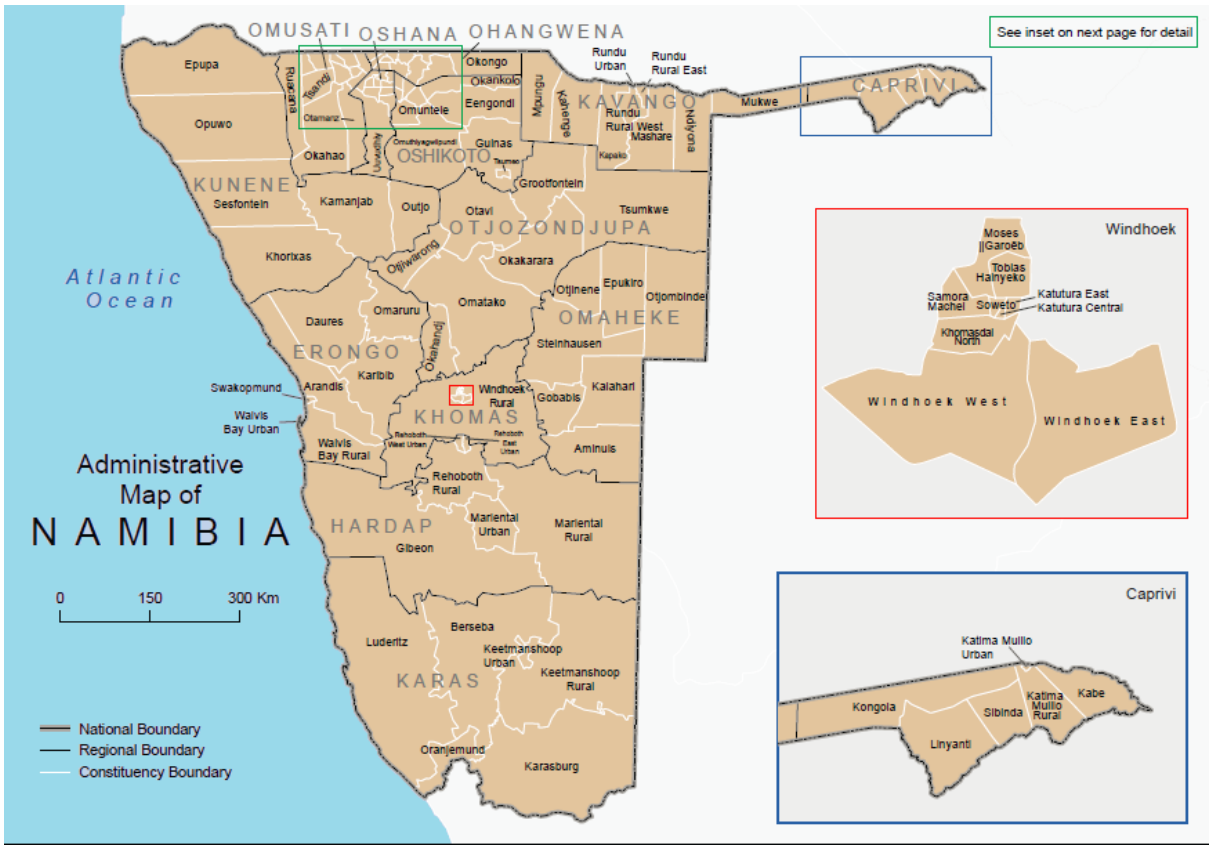
APPENDIX A: Map of the Health districts in Namibia.



APPENDIX B: Regional boundary map of Namibia.



APPENDIX C: Part A of the constituency map of Namibia.



APPENDIX C: Part B of the constituency map of Namibia.