MAPPING THE HUMAN IMMUNODEFICIENCY VIRUS/ACQUIRED IMMUNODEFICIENCY SYNDROME EPIDEMIC IN NAMIBIA USING BAYESIAN SPATIAL MODELLING

A THESIS SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOSTATISTICS

OF

THE UNIVERSITY OF NAMIBIA

BY

JOB SHIKONGO

201069130

September, 2020

Supervisor: Dr. Petrus Iiyambo (Department of Statistics and Population Studies, UNAM)

ABSTRACT

In order to develop an effective prevention response to the Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) epidemic in a particular area, a deeper understanding of the dynamics of the epidemic is required. Disease mapping is usually used to determine the spatial distribution of a disease prevalence and its risk factors in a particular area in order to device appropriate interventions. Although maps of HIV transmission are generally needed for planning, resources allocation, and monitoring and evaluation, such maps were currently not available in Namibia.

This study developed a spatial model of HIV/AIDS epidemic in Namibia based on Bayesian methods (spatially unstructured and structured random effects) using the 2013 Namibia Demographic and Health Survey data. Furthermore, the study identified socio-economic demographic characteristics and sexual behavior that were associated with HIV/AIDS prevalence in Namibia. Specifically, spatial regression models were fitted using BayesX 3.0.2 to adjust for spatial random effects and non-random effects, and the Moran's *I* statistic was calculated to test for the significance of autocorrelation between neighboring regions to show if they tend to cluster.

The Moran's *I* statistic (0.120) was significant (p-value = 0.003) with a variance of 0.002 which stipulated that values that determine the strength of spatial dependence in neighboring regions tend to cluster. After adjusting for spatial random effects and non-random effects, results shows significant structured spatial effects with posterior mean ranging between (-0.423, 0.759) at regional level and (-0.687, 0.995) at constituency level. The socio-economic, demographic and cultural factors like non - condom use, wealth index (poor, middle, richer), marital status (living with partner) and gender (male) were significant in explaining the HIV prevalence in Namibia. Spatial clustering was observed in Khomas, Erongo and towards the regions in the northern parts of Namibia, namely Oshana and Ohangwena.

The study recommends that the modelling of relative risk (as a function of spatial structure and spatial instructed random effects) in Namibia using Bayesian multi-scale models should be based on census data in order to identify definite spatial structures which would be exceedingly critical for both illustrative as well as policy implementation purposes.

TABLE OF CONTENTS

ABSTRACT i

TABLE OF CONTENTS ii

LIST OF FIGURES v

LIST OF TABLES vii

LIST OF ABBREVIATIONS viii

ACKNOWLEDGEMENTS x

DEDICATION xi

DECLARATIONS xii

CHAPTER 1: INTRODUCTION 1

1.1. Background of the study 1

1.2. Statement of the problem 4

1.3. Research objectives 6

1.4. Significance of the study 6

1.5. Limitation of the study 7

1.6. Delimitation of the study 7

CHAPTER 2: LITERATURE REVIEW 8

2.1. Introduction 8

2.2. Global Overview of HIV Epidemic 8

2.3. Determinants of HIV Epidemic 9

2.4. Demographic Factors 10

2.4.1 Age 10
2.4.2. Sex 11
2.4.3. Marital Status 11
2.5. Socio-economic factors 12
2.5.1. Place of residence 12
2.5.2. Religion 13
2.5.3 Education 13
2.5.4. Social Economic Status (SES)/ Wealth 14
2.6. Disease Mapping 14
2.7. Spatial Analysis 16
2.8. Markov Chain Monte Carlo (MCMC) Methods 17
2.9. The BYM model 18
2.10. Models with R-INLA 19
2.11. Model Selection 20
2.12. Deviance Information Criterion 22
CHAPTER 3: METHODOLOGY 24
3.1. Research Design 24
3.1.1. Background of the Namibia Demographic and Health Survey 24
3.2. Population 25
3.3. Sample 25
3.4. Procedures 26
3.5. Data Analysis 26
3.6. Bayesian models for counts 26

3.8. Research Ethics 32

CHAPTER 4: ANALYSIS AND RESULTS 33

4.1. Introduction 33

4.2. HIV Prevalence 37

4.2.1 HIV prevalence with respect to socio-demographic characteristics and sexual behaviours 38

4.3. Random effect analysis 39

4.4. Spatial Effects 45

4.5. Model Selection 48

4.6. Measures of spatial autocorrelation 53

4.6.1. Moran's I Statistic

54

4.7. Discussion 59

5.1. Conclusions 62

5.2. Recommendations 63

REFERENCES 65

ANNEX 1: R-Codes 72

ANNEX 2: BayesX CODES 76

ANNEX 3: Study site HIV testing 79

ANNEX 4: Authorization to download DHS data 80

LIST OF FIGURES

Figure 1: Surveillance Trends in HIV Prevalence (%) among pregnant women receiving antennal care in Namibia for the years 1992 - 2016. 3

Figure 2: log-HIV prevalence rate per Region 33

Figure 3: log-HIV prevalence rate for those who did not use Condom during last sexual intercourse 33

Figure 4: log-HIV prevalence rate: 15 -19 34

Figure 5: log-HIV prevalence rate: 20 - 24 34

Figure 6: log-HIV prevalence rate:25 - 29 34

Figure 7: log-HIV prevalence rate:30 - 34 34

Figure 8: log-HIV prevalence rate: 35 - 39 34

Figure 9: log-HIV prevalence rate:40 - 44 34

Figure 10: log-HIV prevalence rate:45 - 49 34

Figure 11: log-HIV prevalence rate: 50 - 54 34

Figure 12: log-HIV prevalence rate: 55 - 59 35

Figure 13: log-HIV prevalence rate: 60 - 64 35

Figure 14: log-HIV prevalence rate: Male 35

Figure 15: log-HIV prevalence rate: Female 35

Figure 16: log-HIV prevalence rate: Tertiary education 35

Figure 17: log-HIV prevalence rate: Secondary education 35

Figure 18: log-HIV prevalence rate: Primary education 36

Figure 19: log-HIV prevalence rate: No formal education 36

Figure 20: log-HIV prevalence rate: Wealth Index - Poor 36

Figure 21: log-HIV prevalence rate: Wealth Index - Middle 36

Figure 22: log-HIV prevalence rate: Wealth Index - Richer 37

Figure 23: log-HIV prevalence rate: Wealth Index - Richest 37

Figure 24: log-HIV prevalence rate: Rural 37

Figure 25: log-HIV prevalence Urban 37

Figures 26: (a) Spatially structured effects of HIV prevalence at regional level and (b) spatially unstructured random effects of HIV prevalence at regional level 45

Figures 27: (a) Spatially structured effects of HIV prevalence at constituency level and (b) spatially unstructured random effects of HIV prevalence at constituency level 46

Figures 28: (a) Total spatial unstructured effect of HIV prevalence at regional and (b) constituency levels for model 8 presented earlier in Table 3.1..... 47

Figure 29: Moran scatterplot for the spatially lagged HIV positive cases per constituency 57

Figure 30: LISA at constituency level 58

LIST OF TABLES

Table 3.1: Models fitted using BayesX..... 30

Table 3.2: Study variables..... 31

Table 4.1 : Posterior summary statistics of model 1..... 40

Table 4.2: Posterior summary statistics of model 8 presented earlier in Table 3.1..... 43

Table 4.3: Model fit and predictive accuracy at regional and constituency levels..... 49

Table 4.4: Relative Risk and 95% CrI of factors associated with HIV prevalence at regional level 51

Table 4.5: Relative Risk and 95% CrI of factors associated with HIV prevalence at constituency level..... 52

Table 4.6: Five scenarios of LISA..... 56

LIST OF ABBREVIATIONS

AGYW	Adolescent Girls and Young Women
AIC	Akaike Information Criteria
AIDS	Acquired Immunodeficiency Syndrome
ANC	Antenatal Care
BYM	Besag-York-Molli'e
CAR	Conditional Autoregressive
CDC	Centre for Disease Control
CrI	Credible Interval
DHS	Demographic and Health Survey
DIC	Deviance Information Criterion
GMRF	Gaussian Markov Random Field
GPS	Global Positioning System
HIV	Human Immunodeficiency Virus
ICAR	Intrinsic Conditional Autoregressive
ICF	International Coach Federation
INLA	Integrated Nested Laplace Approximation
LGM	Latent Gaussian Model
LISA	Local Indicators of Spatial Association
MCMC	Markov Chain Monte Carlo
MoHSS	Ministry of Health and Social Services
NDHS	Namibia Demographic and Health Survey
NDP5	National Development Plan Goal 5

NHSS	National HIV Sentinel Survey
NIP	Namibia Institute of Pathology
NSA	Namibia Statistics Agency
NSDI	National Spatial Data Infrastructure
NSF	National Strategic Framework
RR	Relative Risk
SES	Social Economic Status
SMR	Standardized Mortality Ratio
UH	Uncorrelated Heterogeneity
UN	United Nations
UNAIDS	Joint United Nations Programme on HIV/AIDS
WBG	World Bank Group
WHO	World Health Organization

ACKNOWLEDGEMENTS

Thank you Lord for your gift of restoration and eternal life, I am thankful for your love and mercy you showed me throughout the writing of this thesis. I would like to thank the entire Department of Statistics and Population Studies at the University of Namibia for granting me this important opportunity to pursue my studies in the field of Biostatistics. In particular, I would like to thank my supervisor, Dr. Peter Iiyambo for his kind guidance, expertise and patience he showed towards me during the completion of this thesis.

I would acknowledge the support I got from my friends, Tobias Shinyemba, Andreas Shipanga, Alugodhi Emvula and Adolf Haufiku for the words of encouragement and wonderful assistance they provided me when the thoughts of giving up were crossing my mind. These good friends supported me immensely and were always willing to help me. The desire to do more and aspirations come from all of them.

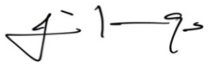
In addition, I would like to thank my parents for their wise counsel and sympathetic ear. You have always been there for me. Finally, there are my friends, who were of great support in deliberating over our problems and findings, as well as providing happy distraction to rest my mind outside of my research at ALNP Hakahana. My family, every day comes bearing its own gift and the first and best victory is to conquer yourselves.

DEDICATION

This thesis is dedicated to the entire Alfred Shikongo and Indongo Iiyambo families. In particular Tuwilika, Twapewa and Tuhanganeni, your unyielding love enriched my soul and inspired me to finish this research and gives me desire to achieve more.

DECLARATIONS

I, Job Shikongo, hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

Job Shikongo		09 September 2020
Name of Student	Signature	Date

CHAPTER 1: INTRODUCTION

1.1. Background of the study

The Human Immunodeficiency Virus (HIV) remains difficult to control due to the complexity of the interactions between human, nature and environmental factors (MoHSS, 2016). Stigma and discrimination remains critical challenges for services uptake and disclosure of HIV status. Not every person is willing to disclose their HIV status but the aggregated data at national level enable researchers to monitor the HIV/AIDS epidemic through high quality analyses for the improvement of social interventions in Namibia. Namibia does not yet have maps of HIV transmission although they are needed for the purpose of planning, monitoring and evaluation, and resource allocation. The fundamental problem of this study was how to analyze HIV prevalence when the geographical information is available and to estimate how the growing social and economic disparities are linked to HIV. Several surveys and studies have been conducted to examine HIV awareness and risk behaviors, and to gain more insight into the social factors that social factors increase vulnerability to infection (Chimoyi & Musenge, 2014; De la Torre, Khan, Eckert, Luna, & Koppenhaver, 2009; L. A. Otwombe, 2013).

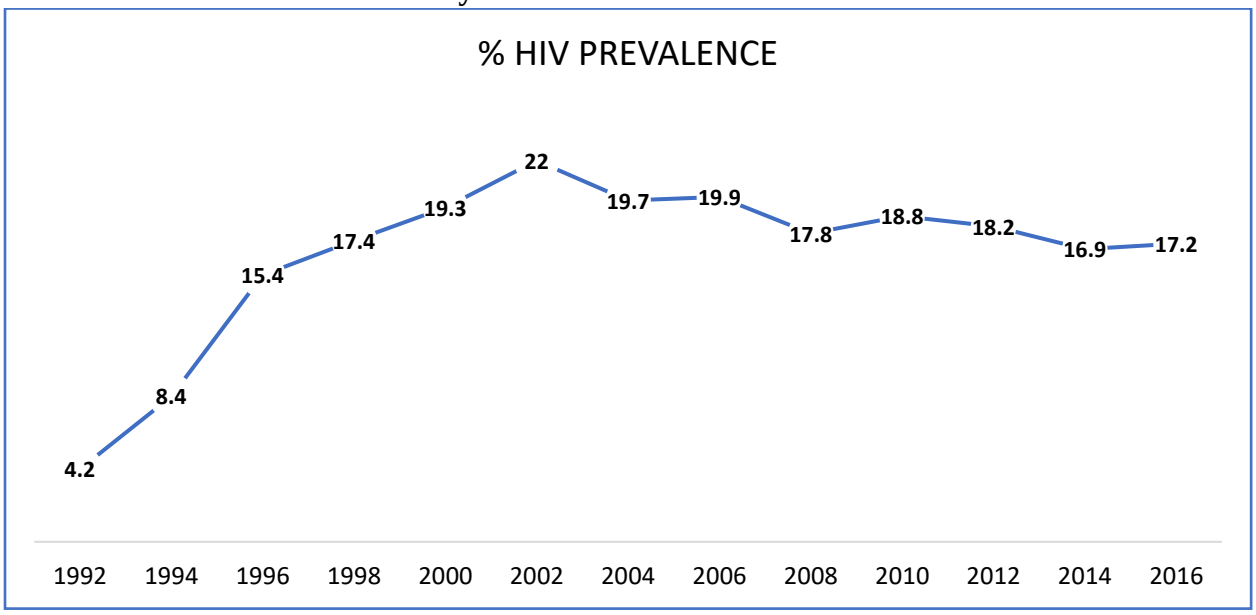
Results from the 2016 National HIV Sentinel Survey (NHSS) suggest that Namibia's epidemic remains in a period of stabilization with slow yet sustained decrease in HIV prevalence among pregnant women since 2002. This is due to concerted efforts by government and various stakeholders. It is further evident, that HIV prevalence trends vary by site, and that the distribution of infection is not uniform across the country. Recent trends show that new infections continue to occur among younger women in Namibia and decision makers need to intensify efforts to prevention-targeted interventions (UNAIDS, 2016).

On geographical comparison a preliminary report by the Center of Disease Control (CDC) (2017) highlighted that among adults aged 15 - 64 years, prevalence of HIV varies geographically across Namibia, ranging from 7.6% in Kunene region to 17.9% and 22.3% in Ohangwena and Zambezi regions, respectively. Regions with higher than national prevalence of 14% tend to have higher HIV prevalence among females than males. The regions with the lowest prevalence were Kunene and Khomas.

Diminishing donor funding and late disbursement of funds to programme throughout the year resulted in delays of procurement, impacting on the effective sustainability of ongoing activities. The lack of effective involvement of some ministries in the Joint United Nations Programme on HIV/AIDS (UNAIDS) response is also still a concern. The Namibian National Strategic Framework (NSF) has not been widely disseminated in different sectors and opportunities for private sector and civil society partnerships have been insufficiently developed (UNAIDS, 2016).

Currently in Namibia the National HIV Sentinel Survey data is used to estimate the HIV prevalence in the general population and for programme purposes, in the absence of a population-based HIV prevalence data that is conducted every five years. The overall HIV prevalence among pregnant women included in the NHSS in Namibia is from 1992 to 2016. The overall HIV prevalence among pregnant women receiving Antenatal Care (ANC) in Namibia was 17.2% during the 2016 NHSS, while in 2014, the overall HIV prevalence was 16.9%. Following a peak of 22% in 2002, HIV prevalence appears to have slowly declined and remained fairly stable during the subsequent years until the year 2016.

Figure 1: Surveillance Trends in HIV Prevalence (%) among pregnant women receiving antennal care in Namibia for the years 1992 - 2016.



Source: Surveillance Report of the 2016 National HIV Sentinel Survey

Maps are often used to spot out areas of a country with the most disease occurrences in order to plan for a proper intervention and targeted distribution of aid to most affected areas (Harris, 2017; Ntirampeba, 2018). They are indeed regarded as useful tools for geographical targeted interventions, monitoring and evaluation of disease burden. However, the construction of such maps is fraught with a number of challenges. One of the setbacks is that these maps are constructed using data that may contain errors. In addition, these maps may inherit the problems due to biased selection methods and the sparse nature of data collected from small geographical areas or ecological fallacy (Ntirampeba, 2018).

Ntirampeba, Neema, and Kazembe (2017) investigated the HIV prevalence using joint spatial modelling of disease risk based on multiple sources. Their findings revealed that health districts and constituencies in the northern part of Namibia were found to be highly associated with HIV infection. Also, the study showed that place of residence, gender, gravida, marital status,

number of children who died, wealth index, education, and condom use were significantly associated with HIV infection in Namibia.

Evidence-based HIV prevention and treatment programme that are cost-effective need to be broadly-diffused globally. Substantial investments must be made in understanding how to implement programme which have clinically-meaningful impact and continuously monitor intervention quality over time (Rotheram-Borus, Davis, & Rezai, 2018). Greater targeting of the right interventions for the right people in the right places at the right times will improve the efficiency and effectiveness of the global response. The HIV/AIDS epidemic will not end without an effective vaccine or cure (Wilson & Taaffe, 2017).

1.2. Statement of the problem

In order to develop an effective prevention response to HIV epidemic, a deeper understanding of its dynamics is required. It has been reported in The Namibian Newspaper (February 12, 2018) that the Global Fund has reduced its financial support to Namibia in order to help other less fortunate countries with bigger public health challenges. As a result, this reduction in financial aid is expected to create a large burden to the health challenges in Namibia. Example of such health challenges are the prevention, treatment and care and support programmes of HIV-AIDS, Malaria, and Tuberculosis. Furthermore, appropriate studies for modelling HIV prevalence might help to strengthen our economy as the monetary meant for HIV epidemic might be diverted to some other equally important tasks. However, less information is available on mapping of diseases in Namibia, a disadvantage in the spatial epidemiology framework which assist Namibia determine the disease risk at regional and local authority levels in order to better deploy limited resources at targeted areas with high risk (Harris, 2017).

The motivation for spatial analysis in epidemiological modelling is based on the notion that people living in a particular household and those live in close proximity share similar exposures to the disease (Musenge, Vounatsou, Collinson, Tollman, & Kahn, 2013). A study by Ntirampeba et al. (2017) on joint spatial modelling of disease risk using multiple sources revealed that health districts and constituencies in the northern part of Namibia were found to be highly associated with HIV infection. Their findings recommended that despite extensive literature on the analyses of determinants of HIV and its geographical spread, research on HIV prevalence using different approaches should be used. Most of the analyses used in the existing literature to investigate the prevalence of HIV/AIDS epidemic were mostly based on univariate methods. Bayesian statistics provides a much more comprehensive depiction of uncertainty in the estimation of the unknown parameters, especially after the confounding effects of nuisance parameters are removed.

Disease mapping has become popular in the field of Statistics as a method to explain the spatial distribution of disease outcomes and as a tool to help design targeted intervention strategies. Most of these models however have been implemented with assumptions that may be limiting or altogether lead to less meaningful results and interpretations. Some of these assumptions included the linearity assumption by allowing the covariates like age to have a non-linear effect on HIV and the stationarity assumption that the relationship between the explanatory variable and the response variables in a regression model are constant and do not change over time across the study region in explaining the spatial distribution of disease outcomes. Studies have shown that the linearity assumption is not necessarily true for all covariates (Okango, Mwambi, & Ngesa, 2016).

1.3. Research objectives

The main objective of this study was to develop a spatial model of HIV/AIDS epidemic in Namibia using the Bayesian spatial modelling techniques.

The specific objectives of this study were:

a) To analyze HIV/AIDS prevalence and identify socio-economic and demographic characteristics as well as sexual behavior that were associated with HIV/AIDS prevalence in Namibia.
b) To fit spatial regression models to the data using BayesX and to identify the best fit model.
c) To determine if there was a significant spatial clustering of HIV/AIDS in Namibia using Moran's *I* statistic.

1.4. Significance of the study

Since Namibia is classified as an upper middle income country by the World Bank in the year 2016 with a Gross Domestic Product (GDP) per capita of around US$5700 at current exchange rates (World Bank, 2018), the financial and technical support the country receive from developing partners (such as UN agencies, World Bank and Global Fund) to fight against the HIV/AIDS epidemic have been declining. Hence there is a need to come up with lasting solutions that will prevent the occurrence of new HIV infections in the country. Considerable increases in funding for HIV/AIDS programmes and a strong commitment by the Government of Namibia present an important opportunity for curtailing the epidemic (De la Torre et al., 2009). The world is moving towards a digital library that documents every inch of our daily lives and Namibia is no exemption. The findings of this study can assist the Ministry of Health and Social Services (MoHSS) to derive health metrics, guide intervention strategies and advance epidemiological understanding using spatial modelling and mapping of diseases.

1.5. Limitation of the study

This study used the 2013 Namibia Demographic and Health Survey (NDHS) data only because the Global Positioning System (GPS) and Community level information is only contained in this dataset while such information is missing in the previous and similar surveys datasets. Furthermore, Demographic and Health Surveys (DHS) include limited measurement of adult health outcomes, particularly chronic disease, as well as limited data collected on men's health and the elderly population.

1.6. Delimitation of the study

According to the 2013 NDHS report, only women and men aged 15 - 64 years in half of the surveyed households were included in the HIV testing. Moreover, some respondents chose not to participate, particularly those individuals who have previously been diagnosed with HIV. Thus the study sample might not be representative of the whole nation.

CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

This chapter presents a review of existing literature on the mapping of HIV epidemic. The literature review focus on how demographics, socio economic and cultural factors contribute to the chance of an individual being infected with HIV virus. Different models of mapping the HIV epidemic are also discussed in this chapter.

2.2. Global Overview of HIV Epidemic

Kilmarx (2009) projected the stability of the global prevalence of HIV with important regional differences in trends and modes of transmission. Prevention and treatment programme have an expanding impact in preventing HIV infection and AIDS related deaths.

About 1 million people died from AIDS-related illnesses in 2016, bringing the total number of people who have died from AIDS-related illnesses since the start of the epidemic to 35.0 million (UNAIDS, 2016). Sub-Saharan Africa remains most severely affected with nearly 1 in every 25 adults (4.2%) living with HIV and accounting for nearly two-thirds of the people living with HIV worldwide (WHO, 2018). In 2016, there were 19.4 million people living with HIV (53%) in Eastern and Southern Africa, 6.1 million (17%) in Western and Central Africa, 5.1 million (14%) in Asia and the Pacific, and 2.1 million (6%) in Western and Central Europe and North America (UNAIDS, 2016). The HIV epidemic does not only affect the health of individuals, but it affects households, communities, and the economic growth of nations. Many of the countries hardest hit by HIV also suffered from other infectious diseases, food insecurity, and other social-economic problems. Despite these challenges, there have been successes and promising signs of scientific understanding of HIV and its prevention and treatment (UNAIDS, 2018). New global efforts have been mounted to address the epidemic, particularly in the last

decade. The number of people newly infected with HIV has declined over the years (UNAIDS, 2018). In addition, the number of people with HIV receiving treatment in resource-poor countries has dramatically increased in the past decade.

According to Zvoushe (2014), African continent is the most affected region in the world where the death of parents due to AIDS is causing many homes to be child-headed. HIV and AIDS have reached pandemic levels in Sub-Saharan Africa, with more than 23 million adults living with the disease.

There are considerable socio-demographic differences in the HIV epidemics in low and middle income countries (young women in Sub-Saharan Africa) compared to high income countries (predominantly gay, bisexual, transgendered youth, especially Black and Latino youth). Researchers and clinicians are designing developmentally-tailored interventions that anticipate youths' engagement with mobile technologies and build on the common features of evidence-based interventions that pre-date the use of antiretroviral therapies (ARV) for prevention and treatment (Rotheram-Borus et al., 2018).

2.3. Determinants of HIV Epidemic

Otwombe (2014) studied the spatial distribution and analysis of factors associated with HIV infection among young people in Eastern Africa. His study found that multiple sexual partnerships with limited condom use and age at first sex debut were associated with HIV prevalence. His findings concluded that the chances of an individual being infected with the HIV virus increased when such individuals engaged in multiple sexual encounters compared to those with a single or no partner.

Similarly, other studies also identified a number of social, economic, demographic and cultural factors that were associated with the risk of HIV infection. These factors included age group,

gender, divorce or widowhood, urban residence, socio-economic conditions and cultural factors. For example, Atilola (2014) investigated the prevalence, spatial patterns and factors associated with HIV infection in Zimbabwe. The findings of this study showed that occupation and employment status were risks associated with infection rates in Zimbabwe. In addition, this study also found an increased risk of HIV infection in urban dwellers compared to their rural counterparts, and that a wide disparity in the risk of HIV infection was known to exist between men and women.

High risk behaviours associated with HIV among Adolescent Girls and Young Women AGYW include early sexual debut, multiple sexual partnerships, limited condom use, intimate partner violence, intergenerational and transactional sex (Mabaso et al., 2018). Furthermore, the socio-demographic factors such as age, marital status, level of education, employment, and place of residence were also found to be associated with risk of HIV among young people.

2.4. Demographic Factors

2.4.1 Age

Age is one of the independent demographic that is often included when modelling effects of demographic characteristics of the respondents on the prevalence of HIV epidemic**.** For example, Otwombe (2014) found that there was no association between age and HIV prevalence in countries such as Ethiopia and Tanzania. However, in countries such as Kenya and Uganda, Otwombe (2014) showed that an increase in age reduced the chances of acquiring HIV/AIDS.

Woldemariame (2013) indicated that most young people were rushing to sex and practicing unsafe sex due to their puberty age. The study further found that the economic status and knowledge of young people about the epidemic also contributed to the prevalence level of HIV

infection. The findings showed that youth were more attracted to money especially if they were poor and have inadequate information about the disease. The study further concluded that the prevalence of the epidemic differed across various age groups. That is, some age groups seem to be more vulnerable to the epidemic than others. Specifically, age group of 15 - 49 years.

2.4.2. Sex

Hollmann (2016) conducted a study on mapping the HIV in Uganda. The study revealed that risk-taking sexual behaviour in both genders was noticeable. In Uganda, 8.2% of all women were infected with HIV, compared to 6.1% of all men. It further stated that women had less access to health care or were less likely to seek health care due to higher discrimination and stigma they faced compared to men. In addition, many women were financially dependent on their husbands or partners, who often refuse the use of condoms. Similar results were also found by Otwombe (2014). That is, females were more likely to be infected with HIV/AIDS across the region compared to their male counterparts.

According to Woldemariame (2013), HIV prevalence level was found to be higher among women than men. The reason for this finding was that women were biologically more susceptible to infection due to larger genital tract surface area, which might be also torn during sexual activity, and this was anticipated lead to higher risk of HIV transmission.

2.4.3. Marital Status

The marital status of individuals can also explain their risk of HIV infection. For example, Atilola (2014) found that unmarried men and women in Zimbabwe and other parts of Sub-Saharan Africa tend to have a higher risk of HIV infection compared to the married counterparts due to their high sexual activity, inconsistency use of condom, and multiple or concurrent sexual partnership.

According to Woldemariame (2013), marital status of an individual affects the transmission of the HIV epidemic through acquiring new sexual partners and less autonomy. As a result, married women were found to have higher levels of sexual activity than unmarried peers. Woldemariame (2013) further found that married women had limited ability to negotiate condom use, and had low power to refuse sex with their partner. Thus, the study showed comparatively higher rates of HIV infection in unmarried persons than in married persons, and concluded that the risk of HIV positive among widows and divorced persons was higher than among those who had never married. This means married people are subjected to more risky activities, and fears about contracting HIV are correlated with potential spouse's sexual habits.

2.5. Socio-economic factors

2.5.1. Place of residence

With regard to the place of residence, living in the rural areas provided a protective factor against HIV sero-conversion with significant associations (Otwombe, 2014). According to Hollmann (2016), urban areas reported higher HIV infection rates than their rural counterparts. The study found that people in urban areas engaged in riskier sexual behavior, more people participated in transactional sex, the urban population tend to have multiple and non-regular partners.

Place of residence of individuals may affect the prevalence rate of the epidemic through socio-cultural and socio-economic variation. Furthermore, rural areas are of no exemption as Woldemariame (2013) study on the factors determining the prevalence of HIV/AIDS in Ethiopia found that the level of knowledge about HIV/AIDS was lower in rural areas compared to urban areas. High HIV infection rates were exhibited in urban areas, but the transmission of the epidemic was frequent among the poor, hence increasingly also in rural areas. The study

revealed that HIV rate was found to be high in rural residents reporting problems of using condoms due to their availability and religious beliefs.

2.5.2. Religion

Religion has been found to have an influence towards HIV infection. According to Otwombe (2014), there exists a relationship between the risk of HIV infection and religion. For example, the religious beliefs among some Muslims and Pentecostal Christians led to low HIV prevalence rates. The fact that HIV/AIDS has generated since the very beginnings of the epidemic a considerable amount of theological work, in Africa as elsewhere in the world, is probably not alien to the gradual change of minds of Christian leaders (Denis, 2013).

2.5.3 Education

When mapping HIV epidemic, education of the participant was also found to be often a significant factor. For example, Hollmann (2016) found that education shaped the behaviors and knowledge of young adults and children in Uganda, and hence it influenced their risk of infection with HIV. The study further concluded that higher educational attainment decreased a risk for HIV infection.

Woldemariame (2013**)** found that the prevalence of HIV/AIDS varied among individuals who have different educational attainments**.** The rate of HIV infection was considerably lower among individuals with tertiary and secondary highest educational qualifications as compared to those who had higher educational attainment. These results showed that better educated individuals had a tendency of preventing the diseases by applying safety methods.

2.5.4. Social Economic Status (SES)/Wealth

Many studies found a relationship between SES and HIV prevalence. The chances of being HIV positive increased for participants in the low SES compared to participants in the high SES category (Otwombe, 2014).

Similarly, Hollmann (2016) investigated the relationship between wealth and HIV/AIDS prevalence. The results showed that AIDS epidemic was commonly perceived as a disease of the poor. They further indicated that often the economically disadvantaged participants had poorer nutrition levels that led to an instable immune system, and subsequently to a higher chance of infection with diseases, including HIV/AIDS. Furthermore, the results showed that wealthy people tend to live longer and hence had a higher lifetime risk of getting infected with HIV and they were more likely to have a higher amount of changing partners due to their greater personal autonomy and spatial mobility. Hollmann (2016) further concluded that the prevalence of transactional and age-disparate sexual relationships were higher among the wealthy.

Employment status was linked to the prevalence of the epidemic through generating income and migration to other places in seeking of employment. This difference in income and migration also had an effect on the likelihood of individuals affected by HIV/AIDS epidemic (Woldemariame, 2013)

2.6. Disease Mapping

Disease mapping goes under a variety of names, some of which are: spatial epidemiology, environmental epidemiology, disease mapping, small area health studies (Lawson, 2013a). However, at the center of these different names are two characteristics. Initial a spatial or geographical dissemination is the concentration thus the overall area of occasions is significant.

This brings the universe of land data frameworks into play, while additionally including spatial insights as a key segment. The subsequent fixing is that of disease and it is the spatial appropriation of infection that is the core interest. Thus, the major issue is the manner by which to break down sickness frequency or pervasiveness when there is availability of geographical data. Sometimes this is called geo-referenced disease data, specifying the labeling of outcomes with spatial tags (Best, Richardson, & Thomson, 2005; Lawson, 2013a).

Over the recent years, there has been much interest in spatial modelling and mapping of disease or mortality rates. Due to inherent sampling variability it is not recommended to inspect crude rates directly, but borrow strength from neighbouring regions to get more reliable region-specific estimates (Riebler, Sørbye, Simpson, & Rue, 2016).

Let E_i denote the number of persons at risk in area i $(i = 1, \dots, n)$. These expected numbers are commonly calculated based on the size and demographic characteristics of the population living in region i (Riebler et al., 2016). Further, let Y_i denote the number of observed cases of HIV in region i. When the disease is non-contagious and rare, it is usually reasonable to assume that

$$Y_i|\theta_i \sim Poisson(E_i\theta_i) \tag{2.1}$$

where $\theta_i, i = 1, \dots, n,$ denotes the underlying true area-specific relative risk.

In the conventional generalized linear regression models applied to spatial data, many studies have assumed stationarity in that the same stimulus of a disease predictor provokes the same response in all parts of the study region (Okango et al., 2016). In general multistage modelling was found to be useful by Aregay, Lawson, Faes, and Kirby (2017) and Pereira, Turkman, and Correia (2017) in addressing the modifiable area unit problem when scaling problem is an issue

because it takes into account the scaling through a shared random effect. Hence, for public health applications, it would be useful to jointly model the risk at different levels to obtain more accurate risk estimates for planning purpose.

2.7. Spatial Analysis

Spatial analysis may be useful in understanding the spread of HIV/AIDS infection by identifying factors that lead to the spread of the disease, describe the geographical distribution and identify significant geo-spots of HIV/AIDS . According to Otwombe (2013), spatial epidemiology was founded on the premise that individuals who lived in close proximities were generally exposed to similar factors which were likely to affect detected outcomes. Bayesian geo-statistical methods are increasingly being utilized in spatial analysis, disease mapping and consequently, decision-making. Their flexibility enables them to integrate spatial correlation and modelling of fixed and random variables. Furthermore, analyzing spatially-referenced data requires researchers to estimate the effect of factors on diseases specific to different regions while accounting for spatial autocorrelation of the data (Best et al., 2005). This challenge can be overcome by using Bayesian-based conditional autoregressive modelling such as spatial clustering, spatial smoothing and spatial autocorrelation which incorporates the spatial structure of the data into the analysis for small area variations (Khana, Rossen, Hedegaard, & Warner, 2018).

Bayesian Spatial analysis has been proven to be powerful in disease mapping (Aregay et al., 2017; Gelaw, Magalhães, Assefa, & Williams, 2019; Liu & Sharma, 2018; Okango et al., 2016; Pereira et al., 2017; Srinivasan & Venkatesan, 2015). For a map divided into i $(i = l, 2, .., n)$ regions, let Y_i denotes the observed count of HIV positive cases in the region i and E_i is the

expected count in the ith region. If Y_i follows the Poisson model with θ_i representing the relative risk in the ith region, then the relative risk is a measure of how much of a risk factor influences the risk of a specified disease (Venkatesan & Srinivasan, 2008).

To provide an estimate of the 'at-risk' population at spatial locations, it is necessary to first pick a measure which represents the intensity of cases 'expected' at such locations. Let that measure be represented by X_i. Two possibilities can be explored. Firstly, rates for the case disease can be obtained from either the whole study window, or a larger enclosing region. Usually, these rates are available only aggregated into larger regions (e.g. census tracts). The rates are obtained for a range of subpopulation categories which are thought to affect the case disease incidence (Lawson, 2013b). Secondly, for diseases which have uncertain etiology, it could be possible that factors underlying the incidence of the disease have a spatial distribution that is spatially dependent and hence the disease incidence could relate to unobserved genetically linked subjects out with the observation region.

2.8. Markov Chain Monte Carlo (MCMC) Methods

As is often in disease mapping, realistic models for maps have two or more levels and the resulting complexity of the posterior distribution of the parameters requires the use of sampling algorithms. In addition, the flexible modelling of disease could require switching between a varieties of relatively complex models. In this case, it is convenient to have an efficient and flexible posterior sampling method which could be applied across a variety of models. Efficient algorithms for this purpose were developed within the fields of physics and image processing to handle large scale problems in estimation (Lawson, 2013a).

A powerful feature of MCMC and the Bayesian approach is that all inference is based on the joint posterior distribution (Lawson, 2013a). The Bayesian approach to parameter estimation, on the other hand, treats parameters as probability distributions, naturally encompassing the estimate's uncertainty. Because the estimated uncertainty of parameters factors into the complexity of the underlying model, Bayesian analysis is also particularly well-suited to aiding model selection between models of differing complexities (Annis, Miller, & Palmeri, 2017). This means that, instead of simulating independent values from the posterior distribution, a sample is drawn by running a Markov chain (which is essentially a collection of dependent random variables), whose stationary distribution is the posterior density. Then, the simulated sample can be used to compute posterior summaries of interest such as mean, quantiles, and tail probabilities, as described by (Blangiardo & Cameletti, 2015; Congdon, 2014).

2.9. The BYM model

The Besag-York-Molli´e (BYM) model only assumes a spatially structured component and cannot take the limiting form that allows for no spatially structured variability. Hence, unstructured random error or pure over dispersion within area, is modelled as spatial correlation, giving misleading parameter estimates (Ntirampeba, 2018). 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 (Harris, 2017). The non-spatial effects usually arise through the variation among the populations at risk due to omitted covariates (Tu & Greenwood, 2012). 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. The Besag, York and Mollie model for Poisson distributed case counts is presented as follows:

$$Y_i \sim Poisson(O_i\lambda_i)$$

$$log(\mu_i) = X_i\beta + U_i \quad (2.2)$$

$$U_i \sim BYM(\sigma_{2,1}, \sigma_{2,2})$$

where Y_i is the response variable for region i. O_i is the 'baseline' expected count, which is specified in formula on the log scale with $log(O_i)$ an offset variable. The X_i are covariates, U_i is a spatial random effect, with a spatially structured variance parameter $\sigma_{2,1}$ and a spatially independent variance $\sigma_{2,2}$.

2.10. Models with R-INLA

The integrated Laplace approximation (INLA) methodology is a deterministic approach to approximate Bayesian inference for latent Gaussian models (LGMs). In most cases INLA is both faster and more accurate than MCMC alternatives for LGMs (Wang, Ryan, & Faraway, 2018). The INLA R package (see www.r-inla.org) can be used for quick and reliable Bayesian inference in practical applications. Recent applications of INLA can be found in Wang et al. (2018). There are three key components required by INLA: the LGM framework, a Gaussian Markov random field (GMRF) and the Laplace approximation.

The theory behind the INLA method is not easy at the first attempt. Even so, it is worth making the effort to understand LGMs and GMRFs as this allows us to distinguish which models can be attempted with INLA from those for which INLA is impossible or just impractical (Wang et al., 2018). There are several advantages to the use of Bayesian approach. For example, the specification of prior distributions allows the formal inclusion of information that can be obtained through previous studies or from expert opinion. In addition, the (posterior) probability that a parameter does/does not exceed a certain threshold is easily obtained from the posterior distribution, providing a more intuitive and interpretable quantity than a frequentist p-value.

A major advantage of INLA is that it returns accurate parameter estimates in short computational time. Additionally, the deviance information criterion is provided for Bayesian model choice (Schrödle & Held, 2011). In addition, within the Bayesian approach, it is easy to specify a hierarchical structure on the data and/or parameters, which presents the added benefit of making prediction for new observations and missing data imputation relatively straightforward (Blangiardo, Cameletti, Baio, & Rue, 2013).

Even for the models which can be fit with a little patience, the cost of experimenting with multiple models is prohibitive. This is an obstacle to the modern style of data analysis which considers many possible models. We must recognize the tremendous success of MCMC methods but there are two main drawbacks. Firstly, such methods are difficult to use, and secondly they are slow during computation (Wang et al., 2018).

Consequently, Integrated Nested Laplace Approximation (INLA), which is an analytic approximation based on the Laplace method, has been recently developed as an alternative to MCMC (Ntirampeba, 2018). The approximate posterior marginals obtained from the INLA procedure can then be used to compute summary statistics of interest, such as posterior means, variances and quantiles. As a by-product of the main computations, INLA also computes other quantities of interest like deviance information criterion (DIC), marginal likelihoods, etc., which are useful to compare and validate models (Blangiardo & Cameletti, 2015).

2.11. Model Selection

There has been a long and continuing debate about whether the issue of selecting a model as a basis for inferences is amenable to a strict mathematical analysis using, for example, a decision theoretic paradigm (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). Furthermore, although it is useful to have measures of fit and complexity, and to combine them into overall

criteria that have some theoretical justification, Spiegelhalter et al. (2002) think that an overformal approach to model selection is inappropriate since so many other features of a model should be taken into account before using it as a basis for reporting inferences, for example the robustness of its conclusions and its inherent plausibility (Congdon, 2014; Spiegelhalter et al., 2002).

Spiegelhalter et al. (2002), suggested a measure of 'effective number of parameters'

$$pD = \boldsymbol{E}_{\theta|y}[-2\ log\{p(y|\theta)\}] + 2\ log\ [p\big(y\big|\tilde{\theta}(y)\big\}]. \tag{2.3}$$

Let $\theta = \boldsymbol{E}[\theta|y]$, then pD = 'posterior mean deviance − deviance of posterior means'. In normal linear hierarchical models:

$$pD = tr\big\{\ \hat{I}(\bar{\theta})\boldsymbol{V}(\theta|y)\big\}, \tag{2.4}$$

where $\boldsymbol{V}(\theta|y)$ is the posterior covariance matrix, $\hat{I}(\bar{\theta})$ is the observed Fisher information evaluated at the posterior mean, and the maximum likelihood estimator of the inverse covariance matrix. In general, the justification of the trace of the covariance matrix depends on the approximate normality of the posterior distribution.

If there is 'vague' prior information $\overline{\theta} \approx \hat{\theta}$, and so that $D(\theta) \approx D(\theta) + \chi_k^2$; then

$$pD\ \approx E[\chi_k^2] = k, \tag{2.5}$$

where k is the degrees of freedom. Thus, by analogy to the Akaike Information Criteria (AIC), a deviance information criterion (DIC) was proposed by Best et al. (2005) as a classical estimate of fit, plus twice the effective number of parameters (DIC='goodness of fit'+'complexity'), as follows:

$$\text{DIC} = D(\bar{\theta}) + 2pD \tag{2.6}$$

$$= \overline{D} + 2pD$$

whereby pD was derived from using an information theoretic argument to measure for the effective number of parameters in a model as the difference between the posterior mean of the deviance and the deviance at the posterior means of the parameters of interest, as shown in equation (2.6) above.

2.12. Deviance Information Criterion

DIC can also be considered as a Bayesian measure of model fit or model adequacy, penalized by an additional complexity term pD (Spiegelhalter et al., 2002). Good model assumption argues for the use of DIC in comparing models that have already been shown to be adequate candidates for explaining the observations. DIC is currently the recommended calculation on the basis of several different estimators, with a preference for posterior means based on parameterizations obeying approximate likelihood normality. By the rule of thumb, the best model is one with the smallest DIC (Ntirampeba, 2018).

Furthemore, suggestions have a similar 'information theoretic' background to frequentist measures of model complexity and criteria for model comparison but are based on expectations with respect to parameters in place of sampling expectations. DIC can thus be viewed as a Bayesian analogue of AIC, with a similar justification but wider applicability. It is also applicable to any class of model, involving negligible additional analytic work or Monte Carlo sampling and appears to performreasonably across a range of examples. However, the pD and DIC deserve further investigation as tools for model assessment and comparison (Spiegelhalter et al., 2002).

In conclusion of this chapter, Spiegelhalter, Best, Carlin, and Van Der Linde (2014) reviewed their 2002 DIC (Spiegelhalter et al., 2002) and concluded that its applicability has remained relevant despite of its problems, for example it is assumed that the specified parametric family

of probability distributions that generate future observations encompasses the true model. This assumption does not always hold, and it is desirable to consider model assessment procedures in that scenario. Nevertheless, DIC has stimulated rich developments, and eagerly awaits routine implementation of worthy alternatives (Spiegelhalter et al., 2014).

CHAPTER 3: METHODOLOGY

3.1. Research Design

This is a cross-sectional study design, using secondary data obtained from the 2013 NDHS dataset. The sampling frame used for the 2013 NDHS was the preliminary frame of the 2011 Namibia Population and Housing Census. The dataset is readily available online (www.measuredhs.com).

3.1.1. Background of the Namibia Demographic and Health Survey

The 2013 Namibia Demographic and Health Survey is the fourth nationally representative, comprehensive Demographic and Health Surveys (DHS) conducted in Namibia. The 2013 NDHS was implemented by the Ministry of Health and Social Services in collaboration with the Namibia Statistics Agency and the National Institute of Pathology (NIP). Technical support was provided by International Coach Federation (ICF), with financial support from the Government of the Republic of Namibia, the United States Agency for International Development, and the Global Fund (MoHSS & ICF-International, 2014).

The DHS surveys were designed to collect data on fertility, family planning, and maternal and child health; assist countries in monitoring changes in population, health, and nutrition; and provide an international database that can be used by researchers investigating topics related to population, health, and nutrition.

The overall objective of the survey was to provide demographic, socioeconomic, and health data necessary for policymaking, planning, monitoring, and evaluation of national health and population programme. In addition, the survey measured the prevalence of anaemia, HIV, high blood glucose, and high blood pressure among adult women and men; assessed the prevalence

of anaemia among children age 6 - 59 months; and collected anthropometric measurements to assess the nutritional status of women, men, and children.

A long-term objective of the survey was to strengthen the technical capacity of local organizations to plan, conduct, process and analyze data from complex national population and health surveys . At the global level, the 2013 NDHS data are comparable with those from a number of DHS surveys conducted in other developing countries.

3.2. Population

The study population consist of participants living with HIV/AIDS in Namibia based on the information in the 2013 NDHS.

3.3. Sample

This study used secondary data and applied the two-stage stratified cluster sampling method to select the sample. The sample comprised of respondents aged between 15 - 64 years both women and men that tested positive for HIV. The sampling frame used for the 2013 NDHS was the preliminary frame of the 2011 Namibia Population and Housing Census (NSA, 2011). The sampling frame was a complete list of all enumeration areas (EAs) covering the whole country. In general, an EA is defined as a geographical area covering an adequate number of households to serve as a counting unit for the population census (NSA, 2011). In rural areas, an EA is a natural village, part of a large village, or a group of small villages while in urban areas, an EA is usually a city block. The 2011 population census also produced a digitized map for each of the EAs that served as the means of identifying these areas (NSA, 2011).

3.4. Procedures

The thirteen regions of Namibia were stratified into 26 sampling strata (13 rural strata and 13 urban strata). Samples were selected independently in every stratum, with a predetermined number of Enumeration Areas selected.

3.5. Data Analysis

Initially the socio-demographic characteristics of the study population were described with unit of measurement for continuous variables and levels or values for categorical variables. Prevalence of HIV infection by proposed socio-demographic and behavioral variables were also carried out. Moran's, I coefficients (Elhorst, 2010) were calculated to investigate the presence of spatial autocorrelation. Furthermore, Multi-scale models were applied as a natural model for disease mapping based on aggregation of underlying individual level risks as guided by Best et al. (2005); Lawson (2013a) and Aregay et al. (2017). The developed models included variables of interest (covariates) to make estimation at both regional and constituency levels.

All statistical analyses were carried out using statistical software R version 3.3.1, BayesX 3.0.2 and QGIS software. BayesX is available at http://www.bayesx.org; QGIS at https://qgis.org and R at https://www.r-project.org. These software can be accessed freely.

3.6. Bayesian models for counts

Louie and Kolaczyk (2006) developed a framework for a multi-scale method for disease mapping in spatial epidemiology. They offered a multi-scale extension of the canonical standardized mortality ratio (SMR), consisting of Bayesian posterior-based strategies for both estimation and characterization of uncertainty. As a result, a hierarchy of informative disease and confidence maps can be produced without the need to first try to identify a single

appropriate scale of analysis. Furthermore, Aregay et al. (2017), explored the framework to estimate relative risk in Poisson-based models for count data.

The models proposed by many of the previous studies involve complex statistical and computational techniques that may not always be easily implemented using standard software. Taking into account the need for simpler and more user-friendly methods, Aregay et al. (2017) proposed multi-scale convolution models to obtain smoother risk estimates for multi-scale data. This study follows the proposed models by Aregay et al. (2017) which are presented below:

Suppose y_k is a vector of observed aggregated outcomes $y_{i,k}$ for spatial unit i at the kth scale level, $k = 0, \dots, k; i = 1, \dots, N_k$, where N_k is defined to be the number of units at the kth level and K denotes the number of levels. The notation $ch(i,k)$ denotes the set of spatial subunits at the $k+1$ level uniquely allocated within the ith unit of the kth level. For count data, the aggregation is expressed as $y_{i,k} = \sum_{l \in ch(i,k)} y_{l,k+1}$ for $k = 0, \dots, k$. The aggregated counts at each level were assumed to be conditional distributed according to a Poisson distribution with mean $\mu_{i,k} = e_{i,k}\theta_{i,k}$.

$$y_{i,k}|e_{i,k} \sim Poisson(\mu_{i,k} = e_{i,k}\theta_{i,k}), \tag{3.1}$$

where, $e_{i,k}$ is the expected number of cases in the ith area at scale k and $\theta_{i,k}$ is the relative risk given by

$$\log(\theta_{i,k}) = a_{o,k} + v_{i,k} + u_{i,k} \tag{3.2}$$

where, $a_{o,k}$ is the intercept at scale level k, wheras $v_{i,k}$ and $u_{i,k}$are the spatially unstructured and structured random effects for unit i at level k, respectively. Let $P_k(\theta_k)$ be joint prior distribution of the N_k components of $\theta_k = (\theta_{i,k}, \dots, \theta_{N_k,k})$ for scale k. Then, the posterior distribution at scale k, which is a combination of the prior distribution and the likelihood function was defined as

$$P_k(\theta_k|y_k) \propto L_k(\theta_k|y_k)p_k(\theta_k) \quad (3.3)$$

Let $y_{l,2}, l = 1, \dots, N_2$ be the constituency level counts of disease and $y_{i,2} = \sum_{l \in ch(i,1)} y_{l,2} i = 1, \dots, N_1$, be the ith region level counts of disease obtained by summing the counts at the constituency level; N_1 and N_2 are the number of regions and constituencies, respectively. A joint convolution model at the region and national levels was presented in equation (3.1). The linkage between these two levels was incorporated in the model by including a shared spatially structured country-level random effect, $u_{i,1_{(l \in i)}}$, where $(l \in i)$ denotes region l within the ith country. For example, if country i has m_i regions, each of the region (tested HIV positive) inherits a common characteristic from their country (those who took the HIV test) via the shared random effect $u_{i,1_{(l \in i)}}$ is common for all those who tested HIV positive and belong to the NDHS 2013. For country i and region l within i, the model is given by

$$y_{i,1} \sim Poisson(e_{i,1}\theta_{i,1}),$$

$$\log(\theta_{i,1}) = a_{0,1} + v_{i,1} + u_{i,1}\ , \quad (3.4)$$

$$yl, 2 \sim Poisson\ (e_{l,2}\theta_{l,2}),$$

$$\log(\theta_{i,2}) = a_{0,2} + v_{l,2} + u_{i,1_{(l \in i)}}, \quad (3.5)$$

Note that in this model the spatially correlated random effect is shared from the country-level model to the regional level model. In other words, the spatial association is defined at the coarser (country) level, and assumed to be the same at the finer (regional) level. Here

$$e_{i,1} = p_{i,1}\frac{\sum y_{i,1}}{\sum p_{i,1}} \quad (3.6)$$

$$e_{l,2} = p_{l,2}\frac{\sum y_{l,2}}{\sum p_{l,2}} \quad (3.7)$$

are the expected rates with $p_{i,1}$ and $p_{l,2}$ denoting the population size at the coarser and finer levels, respectively. For this model and for the other models below, we have assumed a flat prior for the intercept parameters, $a_{0,1}$ and $a_{0,2}$. Furthermore, the uncorrelated heterogeneity

(UH) random effects, $v_{i,1}$ and $v_{l,2}$ are assumed to be normally distributed, that is, $v_{i,1} \sim N(0, sd_{v,1}^2)$ and $v_{l,2} \sim N(0, sd_{v,2}^2)$, whereas the correlated heterogeneity random effect, $u_{i,1}$, is assumed to have a conditional autoregressive distribution, which is the most widely used method because of its theoretical properties, computational and interpretation advantages. Furthermore, $sd_{v,1}^2$ and $sd_{v,2}^2$ represents the variances of the random effects, $v_{i,1}$ and $v_{l,2}$ and the measurement error variances at the country and regional levels, respectively. Aregay et al. (2017) and Lawson (2013a), expressed the intrinsic conditional autoregressive (ICAR) structure as

$$u_i | u_{-i,1} \sim N\left(\bar{u}_{\delta_{i,1}}, \frac{sd_{u,1}^2}{n_{\delta_{i,1}}}\right) \tag{3.8}$$

where

$$\bar{u}_{\delta_{i,1}} = \frac{1}{n_{\delta_{i,1}}} \sum_{l \in \hat{\delta}_{l,1}} u_{l,1}$$

Here, $n_{\delta_{i,1}}$ is the cardinality of $\delta_{i,1}$, which denotes the set of labels of the neighbors of regions i and $u_{-i,1}$ is the set of all random effects not including the ith. For the hyperparameters, $sd_{v,2}$, and $sd_{v,1}$ $sd_{u,1}$ a uniform prior distribution, $U(0, 100)$ was considered.

A simplified approach by Pereira et al. (2017) focuses on the application of Bayesian hierarchical models to count data and proportions and considered the h, the chosen link function which generically proposed that the relative risk for equation (3.4) can be assumed as

$$y_{jt} | \mu_{jt} \sim \pi(y_{jt} | \mu_{jt}), \quad j = 1, \ldots, k, \quad t = 1 \tag{3.9}$$

where π is a generic probability mass function and μ_{jt} depend on covariates and on structured and unstructured random factors through an appropriate link function.

Considering that μ_{jt} is the mean of the total HIV positive cases, the spatial structured random effect is assumed to be

$$y_{jt}|\mu_{jt} \sim Poisson(\mu_{jt}), \quad j = 1, \ldots, k, \quad t = 1, \ldots, i$$

Therefore

$$p(y_{jt}|\mu_{jt}) = \mu_{jt}^{y_{jt}} \, exp(-\mu_{jt})/y_{jt}!, \; y_{jt} = 0, 1, 2... \tag{3.10}$$

In this case the link function is the logarithmic function ($log = h$). The regions have different sample dimensions, so the variation of the total HIV positive cases is affected. To remove this effect, an offset term, which is given by the number of individuals in the sample in each region needs to be added to determine if the cases are count or exposure.

A number of models presented in Table 3.1 below were developed to compare the spatially structured and unstructured random effect of HIV/AIDS prevalence at both Regional and Constituency levels. The Deviance Information Criterion (DIC) was used to compare these models. This criterion aimed to achieve a balance between the adequacy of a model and its complexity.

Table 3.1: Models fitted using BayesX

Models
$Model\ 1 = fixed\ effects$
$Model\ 2 = f\ spatial$ at Regional Level
$Model\ 3 = f\ random$ at Regional Level
$Model\ 4 = f\ spatial + f\ random$ at Regional Level
$Model\ 5 = f\ spatial$ at Constituency Level
$Model\ 6 = f\ random$ at Constituency Level
$Model\ 7 = f\ spatial + f\ random$ at Constituency Level
$Model\ 8 = fixed\ effects + f\ spatial + f\ random$ at Constituency and Regional Levels

3.7. Variable Selection

This study considered variables associated with the spatial distribution of HIV and that have been previously associated with the risk of HIV infection (Atilola, 2014; Musenge et al., 2013; Ntirampeba et al., 2017; Okango et al., 2016). The variables were extracted from the NDHS and are presented in Table 3.2 as follows:

Table 3.2: Study variables

Variable	Description (HIV prevalence rate)	Type
HIV	HIV prevalence	Continuous
Sex	1 = Male, 2 = Female	Categorical
Age	Age of individuals	Continuous
Level of Education	0 = No formal education, 1 = primary, 2 = Secondary, 3 = Tertiary	Categorical
Marital Status	0 = Single, 1 = Married, 2 = Living with partner, 3 = Windowed, 4 = Divorced, 5 = Separated	Categorical
Age group	1 = 15 - 19 years, 2 = 20 - 24 years, 3 = 25 - 29 years, 4 = 30 - 34 years, 5 = 35 - 39 years, 6 = 40 - 44 years, 7 = 45 - 49 years, 8 = 50 - 54 years, 9 = 55 - 59 years, 10 = 60 - 64 years	Categorical
Type of Residence	1 = Urban, 2 = Rural	Categorical
Wealth Index	1 = Poorest, 2 = Poorest, 3 = Middle, 4 = Richer, 5 = Richest	Categorical
Condom use	Condom use during last sexual intercourse (0 = No, 1 = Yes)	Categorical

3.8. Research Ethics

Permission to access NDHS HIV dataset was obtained from the Demographic and Health Survey programme (DHS programme) and the ethical clearance letter to conduct the study was granted by the University of Namibia (UNAM) Centre for Postgraduate Studies.

CHAPTER 4: ANALYSIS AND RESULTS

4.1. Introduction

This chapter presents the results and discussions of methodologies used to map HIV prevalence in Namibia.

The maps (Figures 2 through 24) below presents the distribution of Log-HIV prevalence rate per region for socio-economic and demographic characteristics as well as sexual behavior (age, gender, level of education, marital status, religion, wealth index, type of residence and condom use) of the respondents. A typical use of a logarithmic transformation variable was to pull outlying data from a positively skewed distribution closer to the bulk of the data in a quest to make the variable normally distributed. The socio-demographic and behavioral variables were log-transformed in order to meet the normality assumption.

Figure 2: log-HIV prevalence rate per Region

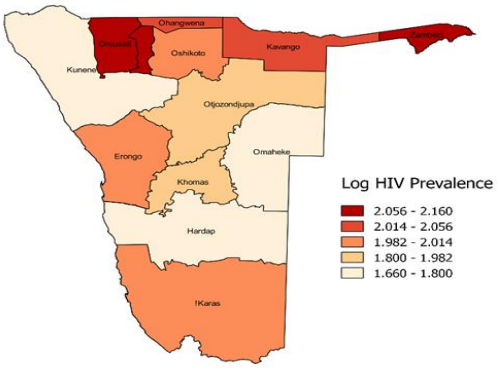


Figure 3: log-HIV prevalence rate for those who did not use Condom during last sexual intercourse

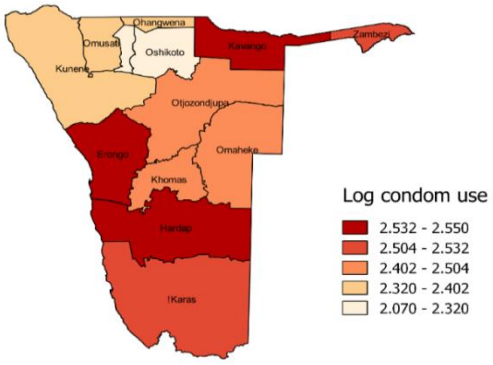


Figure 4: log-HIV prevalence rate: 15 -19

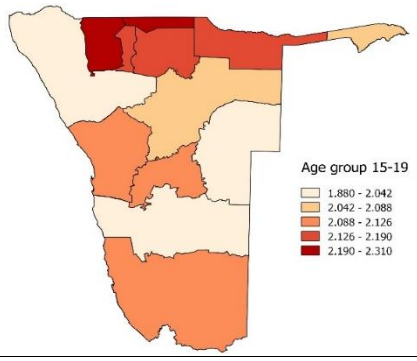


Figure 5: log-HIV prevalence rate: 20 -24

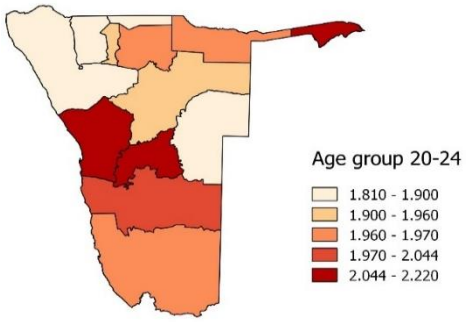


Figure 6: log-HIV prevalence rate:25 -29

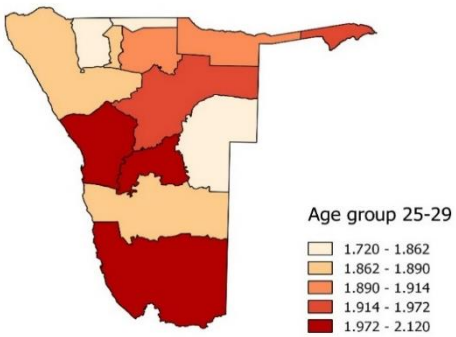


Figure 7: log-HIV prevalence rate:30 -34

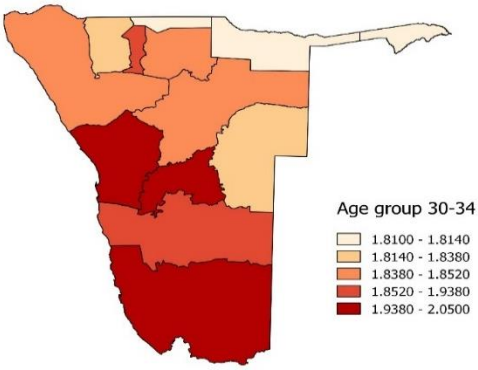


Figure 8: log-HIV prevalence rate: 35 - 39

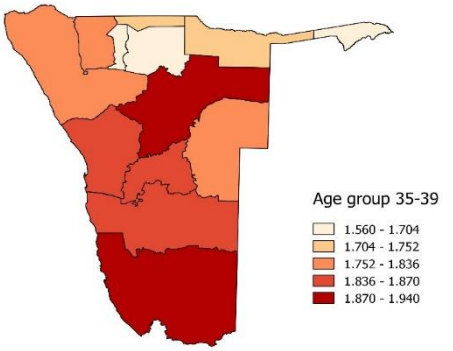


Figure 9: log-HIV prevalence rate:40 -44

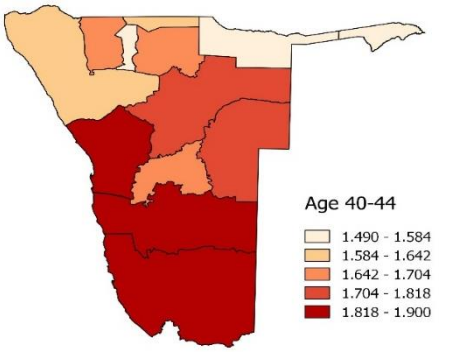


Figure 10: log-HIV prevalence rate:45 - 49

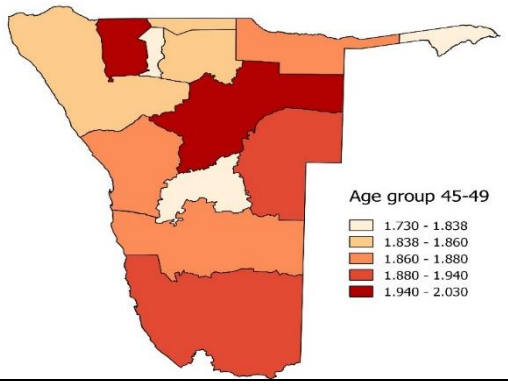


Figure 11: log-HIV prevalence rate: 50 - 54

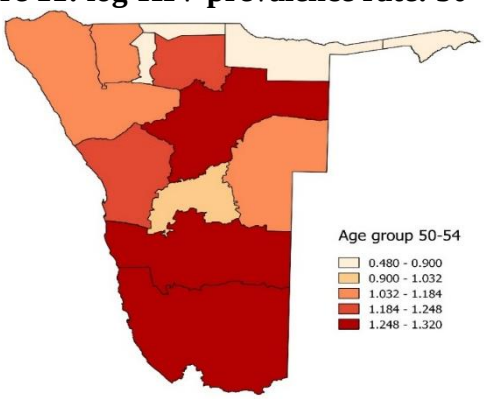


Figure 12: log-HIV prevalence rate: 55 - 59

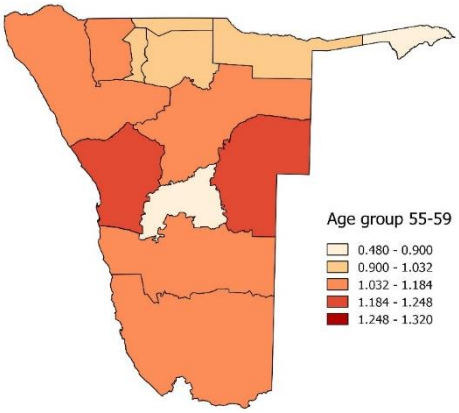


Figure 13: log-HIV prevalence rate: 60 -64

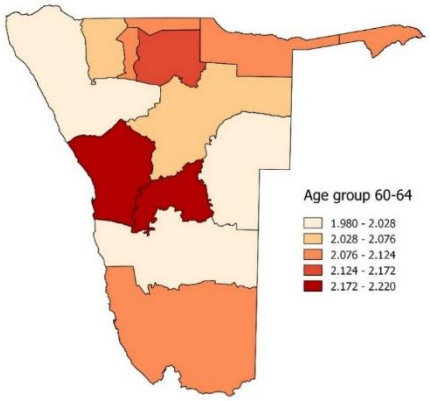


Figure 14: log-HIV prevalence rate: Male

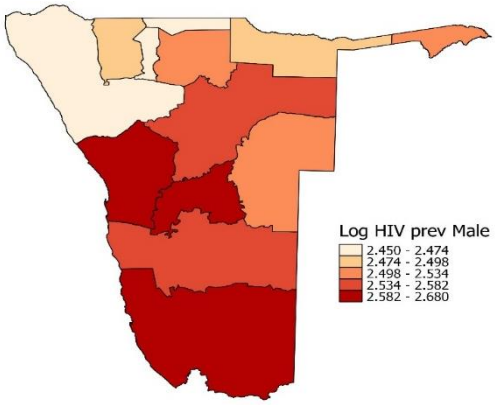
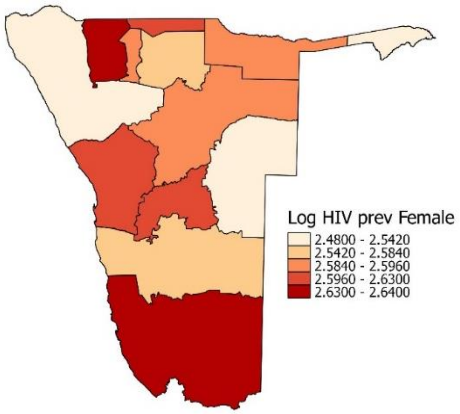


Figure 15: log-HIV prevalence rate: Female



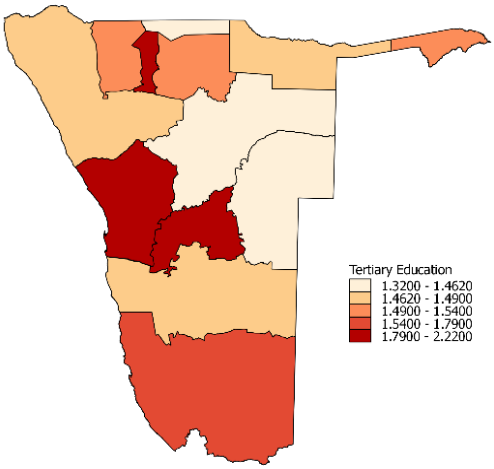


Figure 16: log-HIV prevalence rate: Tertiary education

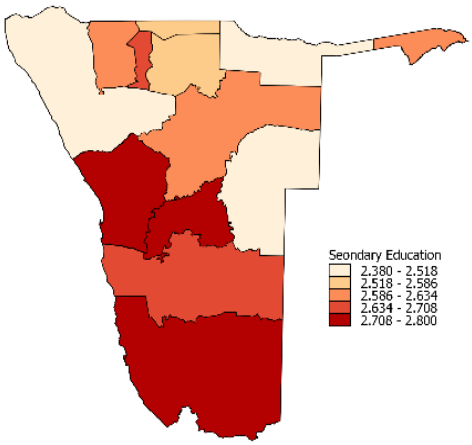


Figure 17: log-HIV prevalence rate: Secondary education

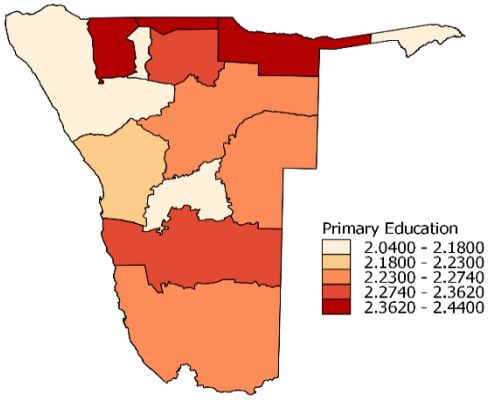


Figure 18: log-HIV prevalence rate: Primary education

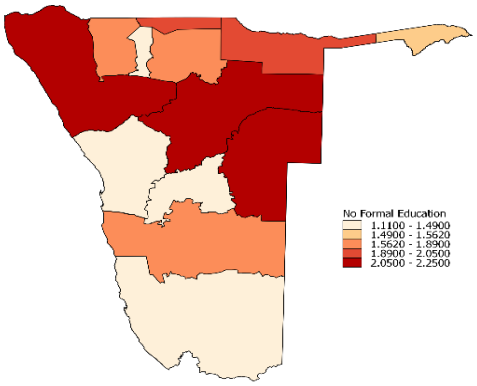


Figure 19: log-HIV prevalence rate: No formal education

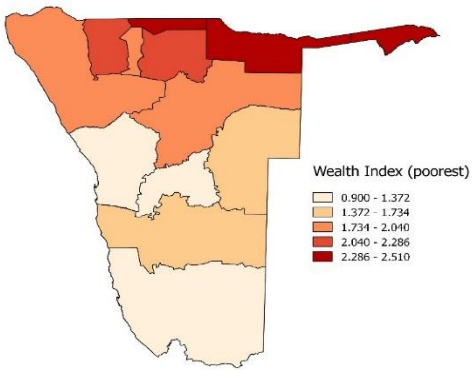


Figure 20: log-HIV prevalence rate: Wealth Index - Poor

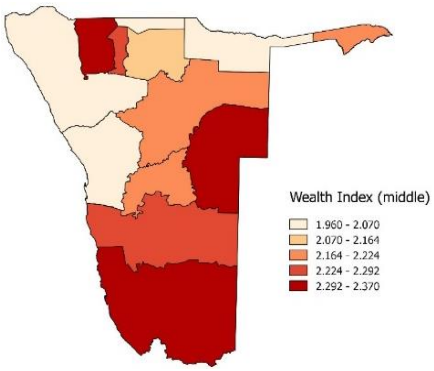


Figure 21: log-HIV prevalence rate: Wealth Index - Middle

Figure 22: log-HIV prevalence rate: Wealth Index - Richer

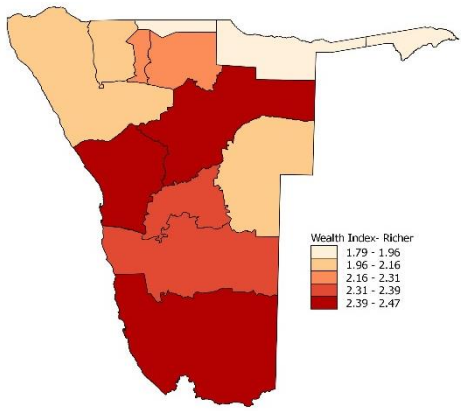


Figure 23: log-HIV prevalence rate: Wealth Index - Richest

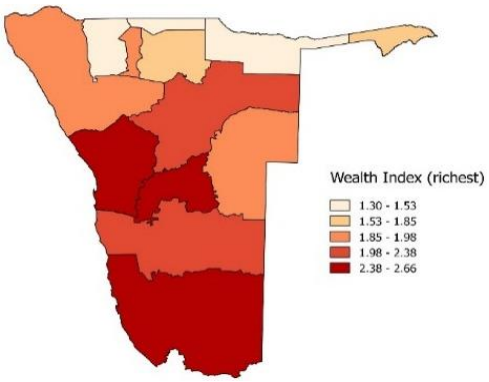


Figure 24: log-HIV prevalence rate: Rural

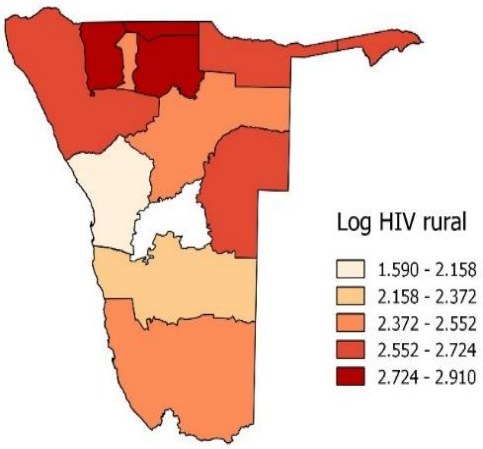
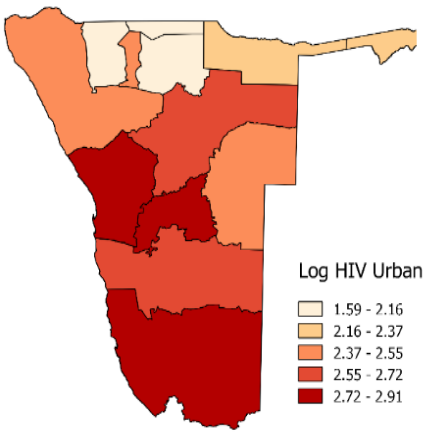


Figure 25: log-HIV prevalence Urban



4.2. HIV Prevalence

The map shown in Figure 1 above, illustrates the distribution of log transformation of HIV prevalence rate per region. Regions in the northern part of Namibia showed the highest prevalence rate. For Zambezi, Omusati, Oshana, Kavango and Ohangwena regions, the result

showed high prevalence rate ranging from 2.014 - 2.160 while corresponding low prevalence rate was observed in Omaheke, Kunene and Hardap regions within the range rate of 1.6 – 1.8.

4.2.1 HIV prevalence with respect to socio-demographic characteristics and sexual behaviours

Figures 2 through 21, presents log-HIV prevalence rates of different socio-demographic characteristics cascaded at regional levels for each category of variables. The variables investigated were condom use during their last sexual intercourse, level of education, sex, age group, type of residence and wealth index. Of the respondents who were HIV positive and did not use a condom during last sexual intercourse, the study found that Kavango, Erongo and Hardap regions had the highest HIV prevalence rates ranging between 2.3 - 2.6 compared to Oshana and Oshikoto region with lowest HIV prevalence rates. Age groups with the highest prevalence rates per region were: 15 – 19 years (for Omusati, Ohangwena, Oshana, Oshikoto, and Kavango); 20 – 24 years (for Zambezi, Erongo, Khomas, and Hardap); 25 – 29 years (for Khomas, Erongo, //Karas, Zambezi, and Otjozondjupa); 30 – 34 years (for Khomas, Erongo, !Karas, and Oshana), 35 – 39 years (for Otjozondjupa, //Karas, Hardap, Khomas, and Erongo), 40 – 44 years (for Erongo, Hardap, Otjozondjupa, //Karas, and Omaheke), 45 - 49 years (for Oshana, Otjozondjupa, and Omaheke), 50 – 54 years (for Otjozondjupa, Hardap, //Karas, Oshikoto, and Erongo), 55 – 59 years (for Erongo and Omaheke) and 60 – 64 years (for Erongo, Khomas and Oshikoto). Overall, the region which recorded the lowest HIV prevalence rate at all age group categories was Kunene.

Erongo, Khomas, //Karas, Hardap and Otjozondjupa regions had the highest HIV prevalence rates for males whereas for females, Oshana, //Karas, Ohangwena, Erongo and Khomas regions recorded the highest prevalence rates. See Figures 13 and 14.

For the respondents with no formal education, the highest HIV prevalence rates (1.96 – 2.250) were observed in Kunene, Otjozondjupa and Omaheke regions compared to //Karas, Khomas and Erongo regions having the lowest HIV prevalence rates (1.110 – 1.449). Similarly, for those with Tertiary Education Erongo and Khomas had the highest HIV prevalence rates. In addition, higher HIV prevalence rates among the respondents who were classified as poorest and were residing in rural areas were observed in Zambezi, Kavango, Ohangwena, Omusati and Oshikoto regions (see Figures 19 and 23). On the contrary, the lowest HIV prevalence rates among the respondents who were classified as poorest and were residing in rural settings were observed in Khomas and Erongo regions (see Figures 19 and 23).

4.3. Random effect analysis

The posterior summary statistics (Relative Risk (RR)), and 95% credible interval (CrI) also known as the Bayesian confidence intervals of the fixed effects that is, the β covariate coefficients, are shown in Table 4.1.

Table 4.1 : Posterior summary statistics of model 1

Variable		RR	95 % Credible Interval	
Constant		0.016	0.007	0.035
Age		1.049	1.033	1.064
Gender	Female (ref)	1.000		
	Male	0.753	0.602	0.948
Marital Status	Single (ref)	1.000		
	Married	0.993	0.685	1.424
	Living with partner	1.752	1.308	2.431
	Windowed	1.107	0.312	3.068
	Divorced	0.689	0.149	2.103
	Separated	1.071	0.57	1.932
Level of Education	No formal Education (ref)	1.000		
	Primary	1.458	1.363	2.259
	Secondary	1.143	0.76	1.76
	Tertiary	0.669	0.338	1.256
Wealth Index	Richer (ref)	1.000		
	Poor	1.724	1.212	2.341
	Middle	1.482	1.082	2.019
	Rich	1.483	1.095	2.009
Religion	No Religion (ref)	1.000		
	Roman Catholic	1.198	0.949	1.515
	Protestant/Anglican	1.015	0.828	1.205
	ELCIN	1.653	1.229	2.293
	Other	0.912	0.719	1.168
Condom Use	Yes (ref)	1.000		
	No	1.641	1.288	2.088
Type of Residence	Urban (ref)	1.000		
	Rural	0.841	0.671	1.064

Note: (ref) represents the reference covariate.

These credible intervals can be used as Bayesian alternatives to the maximum likelihood based 95% confidence intervals, respectively (Hespanhol, Vallio, Costa, & Saragiotto, 2019). The results presented in Table 4.1 do not account for spatial effect and were fitted for model 1 guided by equation (3.2) at 0.05 level of significance. Since equation (3.2) is the Poisson regression, the coefficients of the model were exponentiated and interpreted as RR ($\exp(\theta_{i,k}) = RR$). The RR for a male was 0.735, which represents a very substantial impact on HIV prevalence being 26.5% less in males compared to females. If the 95% credible intervals (CrIs) do not contain 1 then the results were interpreted as being not statistically significant. For males, there was a 95% probability that the true value of the RR would lie between 0.602 and 0.948 which indicated that HIV prevalence rate was less in males and statistically significant. Moreover, regarding the marital status of the respondents, the HIV prevalence was 75.2% more in those who were living with partners with a 95% probability that the true value of RR lies between 1.308 and 2.431, when compared to those were single. Furthermore, for the respondents who were windowed, divorced and separated their CrIs were (0.685 - 1.424), (0.312 - 3.068), and (0.570 - 1.932) respectively. These results indicate that HIV prevalence was either lower or higher among these categories when compared to a respondent being single. The same results of HIV prevalence being either lower or higher among variables was observed between the following coefficients; Level of education (secondary and tertiary education), Religion (Roman Catholics, Protestants/Anglican and other religions).

When comparing wealth index coefficients to the reference category (richest), the risk of HIV prevalence among the respondents increased by 72.4% for those who were poor, 48.2% for those fell in the middle category and 48.3% for those were classified as richer.

The RR for those who did not use a condom during their last sexual intercourse was 1.641 and the 95% CrI for the RR was [1.288 , 2.088] which indicates that the respondents who used a condom during their last sexual intercourse would have lower HIV prevalence compared to those who did not use a condom. The RR associated with living in rural areas was 0.841 which indicated that the risk of HIV prevalence was 84.1% less in rural areas as compared to urban areas. However, these results were contradictory since the CrI were between 0.671 and 1.064 indicating that HIV prevalence could be lower or higher in rural areas when compared to urban areas with at least 95% probability.

The results in Table 4.1 indicate that being a male, living with partner, having a primary education, being poor, belonging to middle class wealth index, being richer, belonging to ELCIN religion and having no used a condom during the last sexual intercourse represented a positive impact on HIV prevalence compared to the respective reference categories (controlling for other significant covariates) with at least 95% probability. Thus, these results were statistically significant at a significance level of 5% in the Bayesian context, since the respective 95% CrIs do not contain the value 1.

Table 4.2: Posterior summary statistics of model 8 presented earlier in Table 3.1

Variable		RR	95% Credible Intervals	
Constant		0.015	0.007	0.030
Age		1.054	1.039	1.070
Gender	Female (ref)	1.000		
	Male	0.754	0.601	0.961
Marital Status	Single (ref)	1.000		
	Married	0.909	0.625	1.262
	Living with partner	1.915	1.400	2.606
	Windowed	0.986	0.269	2.673
	Divorced	0.605	0.137	1.830
	Separated	1.136	0.602	2.034
Level of Education	No formal Education (ref)	1.000		
	Primary	1.269	0.812	2.090
	Secondary	0.915	0.592	1.482
	Tertiary	0.568	0.304	1.024
Wealth Index	Richer (ref)	1.000		
	Poor	1.598	1.144	2.257
	Middle	1.553	1.123	2.122
	Rich	1.605	1.202	2.145
Religion	No Religion	1.000		
	Roman Catholic	1.126	0.878	1.431
	Protestant/Anglican	1.024	0.848	1.215
	ELCIN	1.481	1.080	2.071
	Other	0.897	0.710	1.151
Condom Use	Yes (ref)	1.000		
	No	1.676	1.316	2.137
Type of Residence	Urban (ref)	1.000		
	Rural	0.966	0.698	1.075

Note: (ref) represents the reference covariate.

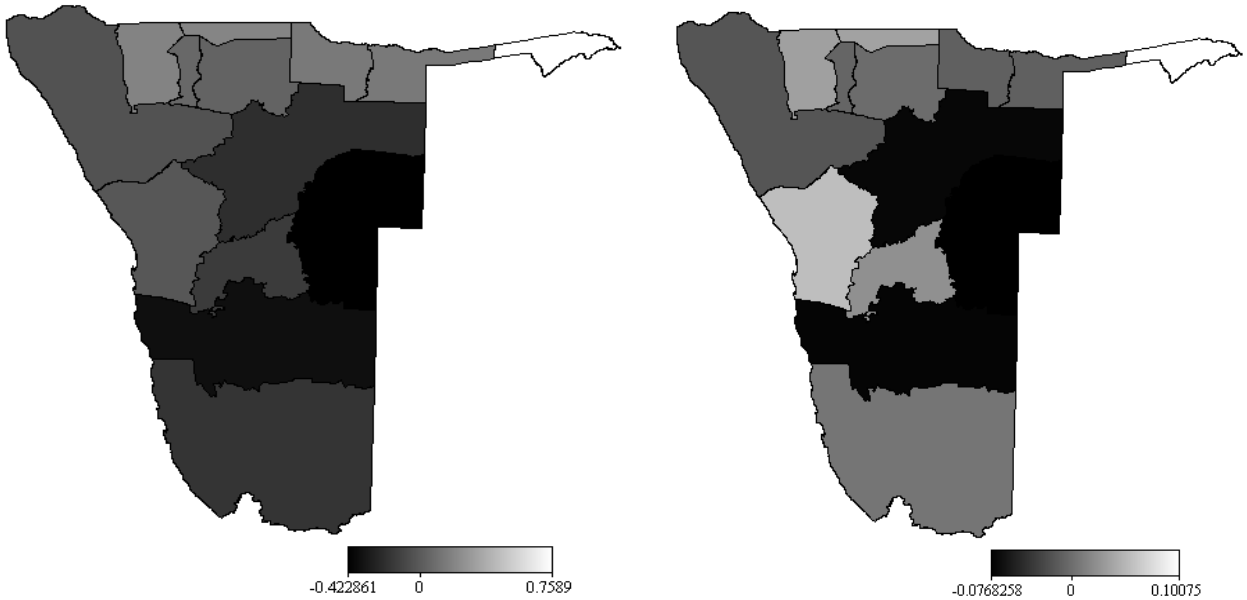
Table 4.2 above presents the fixed effects of RR and 95 % CrI for the HIV prevalence of the socio-economic, demographic and behavioral characteristics of model 8. Model 8 included a spatially random structure effect at both regional and constituency levels. This random effect captured the additional spatial structure to the finer level (equation 3.5), which was not captured by the course level random effect (equation 3.4). In addition, this model introduced the linkage of HIV prevalence between the regions and constituencies and assumed their independence. The respondents who were living with their partners [RR = 1.915; 95% Crl (1.4; 2.606)] and non-condom use during last sexual intercourse [RR = 1.676; 95% CrI (1.316; 2.137)] had a substantial impact on HIV prevalence being high relative to those who used condoms at both regional and constituency levels. Furthermore, by increasing the age of the respondent by one unit [RR = 1.054; 95% CrI (1.039; 1.070)], the study found that HIV prevalence would increase by 5.4% when other covariates are fixed.

HIV prevalence among those who resided in rural areas [RR = 0.966; 95% CrI (0.698; 1.075)] slightly differed by 3.4% from the HIV prevalence of those who resided in urban areas.

Wealth index level also played a critical role on HIV prevalence rate, whereby the poor, middle and richer showed high HIV prevalence when compared to the richest with at least 95% probability. On the level of education, there was uncertainty in determining whether HIV prevalence was high/low among primary education [RR=1.269; 95% Crl (0.812; 2.090)], secondary education [RR=0.915; 95% Crl (0.592; 1.482)] and tertiary education [RR = 0.568; 95 Crl (0.304; 1.024)], when compared to no formal education. The same was observed for Roman Catholics [RR=1.126; 95% CrI (0.878; 1431)], Protestants/Anglicans [RR=0.878; 95% CrI (0.848; 1.215)] and other religion [RR=0.879; 95% CrI (0.710; 1.151)].

4.4. Spatial Effects

Figures 26: (a) Spatially structured effects of HIV prevalence at regional level and (b) spatially unstructured random effects of HIV prevalence at regional level .

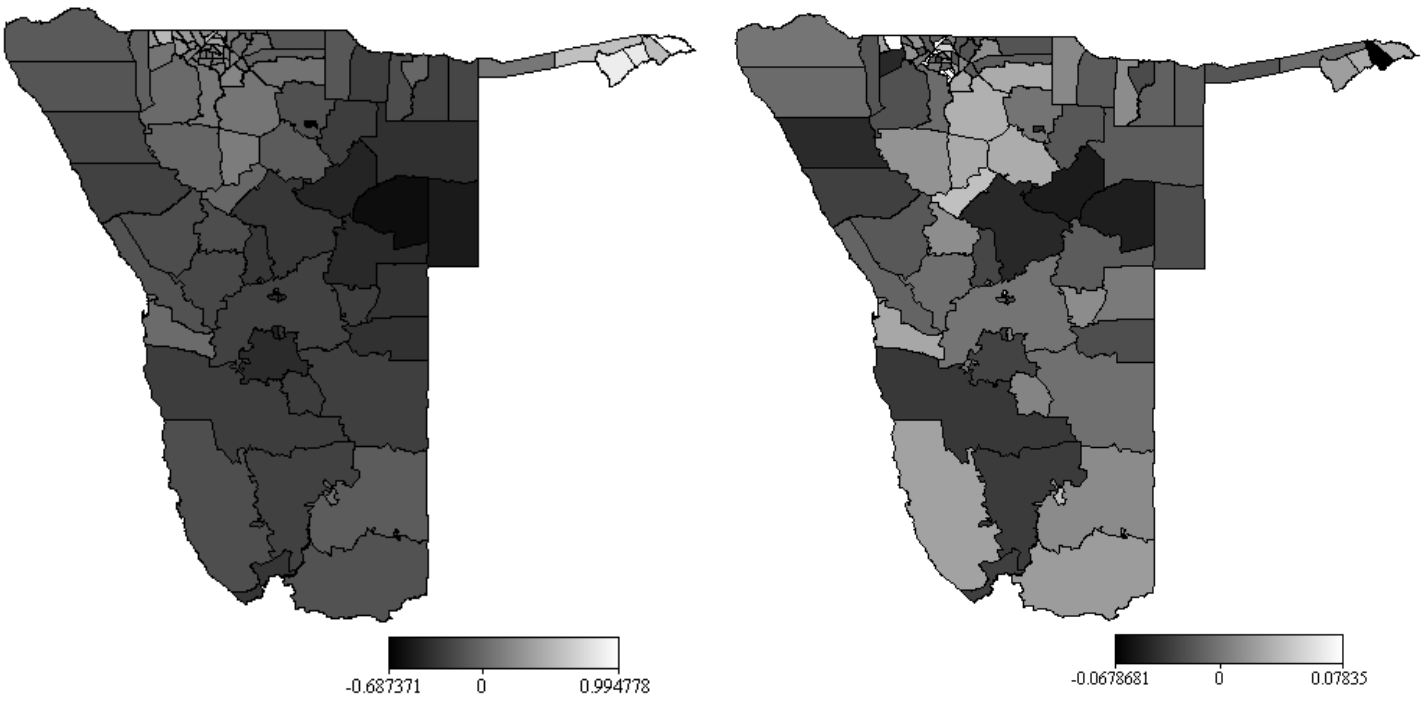


(a) (b)

Figures 26 (a) and (b) above show maps of the spatial structured and unstructured effect of HIV prevalence at regional level respectively, with posterior credible intervals for a nominal level of 95%. Black denotes regions with strictly negative credible intervals (low risk), white denotes regions with strictly positive credible intervals (high risk). The spatial effects stand for unknown influences where these influences may have local or global effects (Habyarimana, Zewotir, Ramroop, & Ayele, 2016). The findings confirmed the evidence of the residual effect on HIV, with the regions in white showing significant positive effects, black negative effects, and light dark non-significant structured spatial effects. The same is interpreted for unstructured spatial effects. The structured model has a posterior mean ranging from -0.434 to 0.759 and unstructured model had posterior mean ranging from -0.077 to 0.101. The significant

factors associated with HIV prevalence per region for spatial structure showed that positive CrI were observed more in Zambezi, Kavango East and Kavango West, Ohangwena, Omusati, Oshana and Oshikoto regions. The regions that showed negative effects were Hardap and Omaheke. This means that HIV prevalence was negatively associated with these regions. For spatially unstructured random effects, Erongo, Omusati, Ohangwena and Khomas regions recorded positive socio-demographic and sexual behavioral characteristics contributing to HIV prevalence rate, while Otjozondjupa, Omaheke and //Karas regions showed significant negative effects.

Figures 27: (a) Spatially structured effects of HIV prevalence at constituency level and (b) spatially unstructured random effects of HIV prevalence at constituency level .



(a) (b)

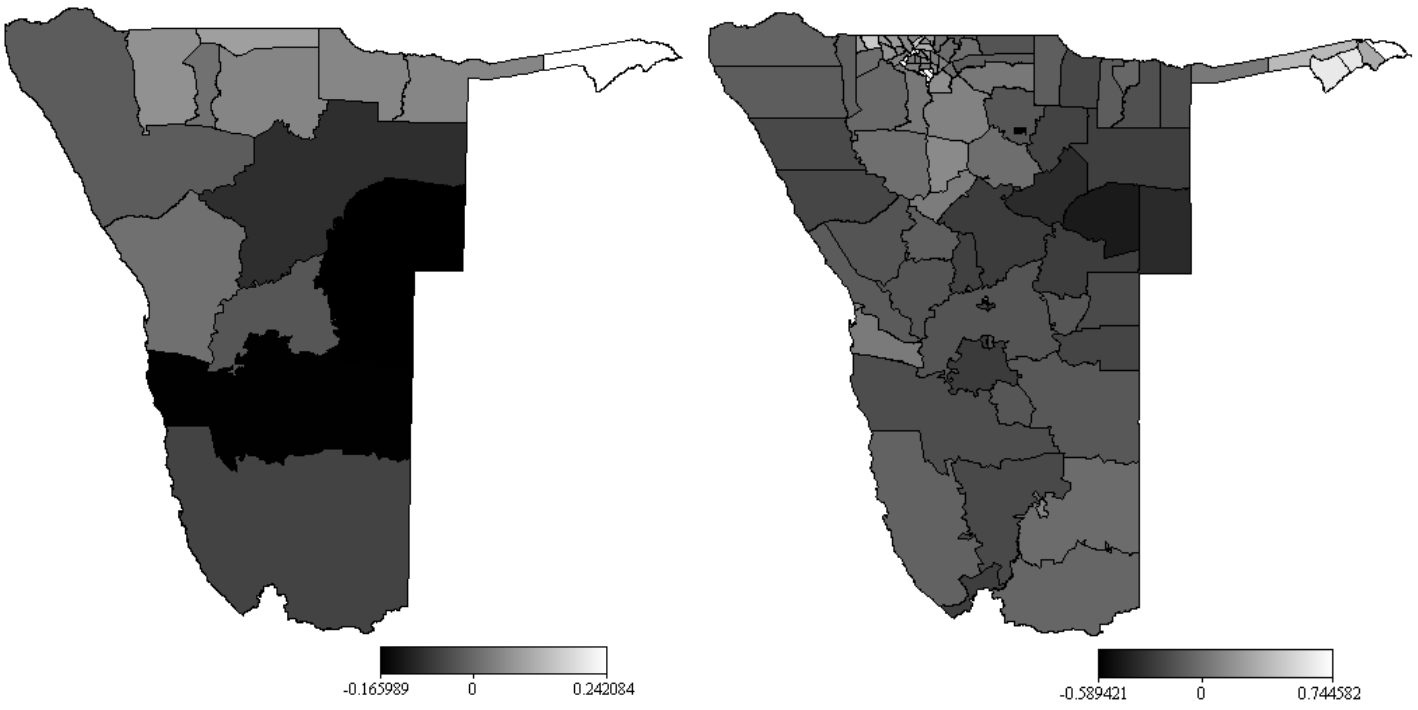
The result illustrated by Figures 27 (a) and (b) assume the spatially structured and unstructured random effect handling correlation between neighboring constituencies, respectively. Thus it is evident that unstructured spatial effect model were the best, when data were simulated from

a Poisson distribution with their posterior mean ranging from -0.687 to 0.995 because there was a presence of spatial random effect.

There was a variation of HIV prevalence among many constituencies in the northern parts of Namibia (specifically in Oshana, Omusati and Ohangwena regions), with positive unstructured spatial effects of HIV prevalence rate when compared to spatial structure effect. The lighter areas on the maps indicate a significant positive relationship between neighboring constituencies with respect to HIV prevalence observed between constituencies in Zambezi region.

Figures 28: (a) Total spatial unstructured effect of HIV prevalence at regional and (b) constituency levels for model 8 presented earlier in Table 3.1

(a) (b)



In addition, the total spatial unstructured effect of HIV prevalence at regional and constituency levels for model 8 are presented in Figure 28 and the fixed effects were presented in Table 4.2 above. The unstructured spatial effect model were best, when data was simulated from a

Poisson distribution with their posterior mean ranging from significant negative random effects to positive random effects.

For Omaheke and Hardap regions, significant negative random effects associated with HIV prevalence were observed with -0.166 posterior mean. Moreover, when these two regions were compared at constituency levels, Hardap region's dark grey colors implied significant low negative random effects associated with HIV prevalence observed in two constituencies (Rehoboth rural and Rehoboth west) and Omaheke region had three constituencies (Epukiro, Otjinene and Otjombinde) with significant negative random effects with posterior mean equal to -0.589. This means that Omaheke and Hardap regions' HIV prevalence had negative random effects because of the impact these constituencies (Rehoboth rural, Rehoboth west, Epukiro, Otjinene and Otjombinde) played respectively, while other constituencies (Daweb, Okorukambe, Kalahari and Aminuis) in these regions had posterior mean of zero, denoting, non-significant random effects were observed in these regions. Furthermore, the whole of Zambezi region had significant positive random effect, but when cascaded at constituency level, only six constituencies (Katima Mulilo Urban, Kabbe North, Kabbe South, Sibbinda, Linyanti and Judea Lyaboloma) showed significant positive random effects.

4.5. Model Selection

DIC is used as a model selection criteria and is currently the recommended calculation on the basis of several different estimators, with a preference for posterior means based on parameterization obeying approximated likelihood normality (Spiegelhalter et al., 2002). The model with the smallest DIC is the best.

Table 4.3: Model fit and predictive accuracy at regional and constituency levels

Model	**Deviance**	$\boldsymbol{pD}$	**DIC**
Model 1 (fixed effects)	1764.95	19.41	1803.77
Model 2 (Region)	1721.90	25.34	1772.58
Model 3 (Region)	1721.05	28.42	1777.91
Model 4 (Region)	1720.97	26.77	1774.51
Model 5 (Constituency)	1680.14	40.47	1761.05
Model 6 (Constituency)	1677.42	50.35	1778.43
Model 7 (Constituency)	1667.67	42.37	1762.40
Model 8 (Region and Constituency)	1680.36	41.62	1763.61

Table 4.3 presents the deviance summaries and predictive accuracy at regional and constituency levels of the fitted models. The models containing spatial random effects (either with or without additional exchangeable effects) both had effective parameters at regional level around 25, and effective parameters at constituency level around 40 to 50, which was concluded as the difference between the posterior mean of the deviance and the parameters.

When comparing the DIC for each model, we first noted that DIC was subjected to Monte Carlo sampling error, since it is a function of stochastic quantities generated under an MCMC sampling scheme. While computing the precise standard errors for the DIC values remains a subject of on-going research, the standard errors for the deviance values were readily obtained and provided a good indication of the accuracy of DIC and pD. In any case, in several runs using different initial values and random-number seeds for this study, the DIC and pD estimates obtained varied by more than 1. According to the rule of thumb, the model with smallest DIC is best fit. Thus, Model 2 which included a fixed and random effect at regional level had the

smallest DIC of 1772.58 and 25 effective parameters was regarded to be the best fit compared to Models 3 and 4.

Similarly, when comparing Models 5, 6 and 7 at constituency levels, Model 5 had the smallest DIC of 1761.05 and therefore was the best fit. Overall, Models 2 and 5 were composed of spatially structured random effects and provided the best fit, with Model 5 providing overall best fit to HIV prevalence data. Furthermore, Model 7 had a DIC of 1763.61 and effective number of parameters, pD= 41, assumed total spatial random effects (unstructured) was also close to best fit models. The spatial effect of the regions and constituencies were split up into a spatially correlated part $fspatial$ and an uncorrelated part $frandom$, as motivated by Belitz, Brezger, Kneib, Lang, and Umlauf (2009). The correlated part was modeled by a (quadratic) pairwise difference penalty, where the neighborhood matrix and possible weights associated with the neighbors were obtained from the map object in BayesX. The uncorrelated part was modelled by an i.i.d. Poisson random effect. The posterior summaries of the model discussed are given in Tables 4.4 and 4.5 below.

Table 4.4: Relative Risk and 95% CrI of factors associated with HIV prevalence at regional level

Variable		Spatial effect RR (95% Crl)	Random Effect RR (95% Crl)
Gender	Female (ref)	1.000	1.000
	Male	0.746 (0.593; 0.954)	0.744 (0.582; 0.944)
Marital Status	Single (ref)	1.000	1.000
	Living with Partner	1.878 (1.390; 2.525)	1.882 (1.371; 2.579)
Wealth Index	Richest (ref)	1.000	1.000
	Poor	1.616 (1.161; 2.198)	1.605 (1.166; 2.210)
	Middle	1.572 (1.141; 2.153)	1.583 (1.170; 2.166)
	Richer	1.613 (1.218; 2.196)	1.615 (1.187; 2.159)
Religion	No Religion (ref)	1.000	1.000
	ELCIN	1.468 (1.049; 2.001)	1.469 (1.081; 1.993)
Condom Use	Yes (ref)	1.000	1.000
	No	1.679 (1.335; 2.127)	1.678 (1.312; 2.140)

Note: (ref) represents the reference covariate.

Table 4.5: Relative Risk and 95% CrI of factors associated with HIV prevalence at constituency level

Variable		Spatial effect RR (95% Crl)	Random Effect RR (95% Crl)
Gender	Female (ref)	1.000	1.000
	Male	0.744 (0.582; 0.944)	0.755 (0.587; 0.946)
Marital Status	Single (ref)	1.000	1.000
	Living with Partner	1.882 (1.371; 2.579)	1.938 (1.401; 2.648)
Level of Education	No formal (ref)	1.000	1.000
	Primary	1.316 (1.212; 2.160)	1.403 (1.252; 2.140)
Wealth Index	Richest (ref)	1.000	1.000
	Poor	1.605 (1.166; 2.210)	1.587 (1.150; 2.215)
	Middle	1.582 (1.170; 2.166)	1.543 (1.123; 2.146)
	Richer	1.615 (1.187; 2.159)	1.613 (1.188; 2.174)
Religion	No Religion (ref)	1.000	1.000
	ELCIN	1.469 (1.081; 1.993)	1.484 (1.050; 2.040)
Condom Use	Yes (ref)	1.000	1.000
	No	1.678 (1.312; 2.140)	1.696 (1.329; 2.142)

Note: (ref) represents the reference covariate.

In summary, Tables 4.4 and 4.5 above, shows the posterior summaries of Models 2 through 7, of factors associated with HIV prevalence at regional (Table 4.4) and constituency (Table 4.5) levels. For spatial effects and random effects, both tables presented similar variables, with Table 4.5 having additional variable (level of education). The results shows that among males, HIV prevalence was significantly less (RR = 0.744) with 95% probability when compared to females in both tables. Further, it was noted that RR among males of spatial effects and random

effects slightly differed at both regional and constituency levels for all the variables in the models.

With respect to condom usage, HIV prevalence for non-condom users during their last sexual intercourse was high with 67.8% (spatial effect) and 69.6% (random effect) compared to those who have used condoms. The RR and their corresponding 95% CrIs presented in Tables 4.4 (at regional level) and 4.5 (at constituency level) were quite similar for both spatial and random effects, which suggests that the random effects and spatial effects only slightly differ. Thus the contradictory explanation of Model 7 being close to a best fit was confirmed by these results.

4.6. Measures of spatial autocorrelation

Many studies involving spatial modelling have often found that a correlation exists between spatial units (Harris, 2017; Paradis, 2018b; Plant, 2012; Shipanga, 2019). This correlation is usually geographical and relates to the basic idea that locations close together in space often have similar values of outcome variables while locations that are far apart tend to have different values (Lawson, 2013a).

In spatial applications, it is further important to distinguish between two basic forms of extra variation (Lawson, 2013b). Firstly, as in the case of a spatial application, a form of independent and spatially uncorrelated extra variation can be assumed. In addition, there could also be correlated heterogeneity. Essentially, this form of extra variation implies that there exists spatial autocorrelation between spatial units. This autocorrelation could arise for a variety of reasons. For example, the disease being modelled could be naturally clustered in its spatial distribution at the scale of observation. Many infectious diseases display such spatial clustering; a number of apparently non-infectious diseases also cluster. Secondly, autocorrelation can be induced in spatial disease patterns by the existence of unobserved

environmental or frailty effects. Hence, the extra variation observed in any application could arise from confounding variables that have not been included in the analysis.

4.6.1. Moran's *I* Statistic

The measure of the strength of spatial similarity between areas of interest can be quantified using spatial autocorrelation. Moran's *I* is a measure of spatial autocorrelation that describe how the values of a variable are related based on the locations where they were measured (Paradis, 2018a). The Moran's *I* statistics developed by (Moran, 1947) and Geary's *c* statistic by (Geary, 1954) test for the null hypothesis that there is zero autocorrelation based on variables measured at interval or ratio levels. The theory of Moran's *I* and Geary's *c* is extensively discussed by Plant (2012).

Quite not so intuitively, the expected value of Moran's *I* statistic under the null hypothesis of no autocorrelation is not equal to zero, but it is given by $I_0 = -1/(n-1)$, where n denotes the number of spatial units. If the observed value of I_0 (denoted $\hat{I}$) is significantly greater than I_0, then values of y (where y denotes the number of HIV positive cases) are positively autocorrelated, whereas if $\hat{I} < I_0$, this indicates that the values of y are negatively autocorrelated way (Paradis, 2018a).

Moran (1947) and Poh-Chin, Fun-Mun, and Ka-Wing (2009) algebraically expressed Moran's *I* statistic as

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j \omega_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} \tag{4.1}$$

where n is the is the number of spatial units (pixels) indexed by i and j, $S_0 = \sum_i \sum_j \omega_{ij}$, which is the aggregate of all spatial weights and ω_{ij} is the spatial weight between locations i and j.

The form of Moran's *I* statistic is similar to that of Pearson's correlation coefficient and gives a score ranging between –1 and 1 (Poh-Chin et al., 2009). A positive value of Moran's *I* statistic means a "hot" spot or that a polygon or point with a high score has other polygons or points with high scores surrounding it. Conversely, an occurrence of a low score indicates a "cold" spot because of low scoring occurrences in the neighborhood. A value of zero for Moran's *I* statistic indicates that nothing can be assumed about the scores of the neighboring polygons or points. A negative value of Moran's *I* statistic indicate a "spatial outlier" or that the scores of neighboring locations would be the opposite of the location of interest. That is, a polygon or point with a low score will have high scoring neighbors, and vice-versa (Poh-Chin et al., 2009).

Although the strength of Moran's coefficient lies in its simplicity (data may are modelled as points or polygons and are assumed to be under the null hypothesis according to the normality assumption or the randomization assumption), its major limitation is the tendency to average local variations in situations of spatial autocorrelation (Plant, 2012). Furthermore, Plant (2012) defined the LISA (Local Indicators of Spatial Association) a function to estimate the local indicators of spatial association. Moreover LISA in R-software can be regarded as the local equivalent of Moran's *I* in showing hot and cold spots (clusters of high and low scores, respectively) as well as spatial outliers (where there is a mixture of high and low scores in neighboring areas) (see Table 4.6). LISA statistics are defined for each cell of an a real data set and serve as an indication of spatial clustering of similar values (Plant, 2012). In addition, however, LISA statistics have the property that their sum, when taken over all locations, equals the value of the corresponding global statistic. LISA statistics therefore provide an indicator of how areas in the region contribute to the value of the global statistic and defined as

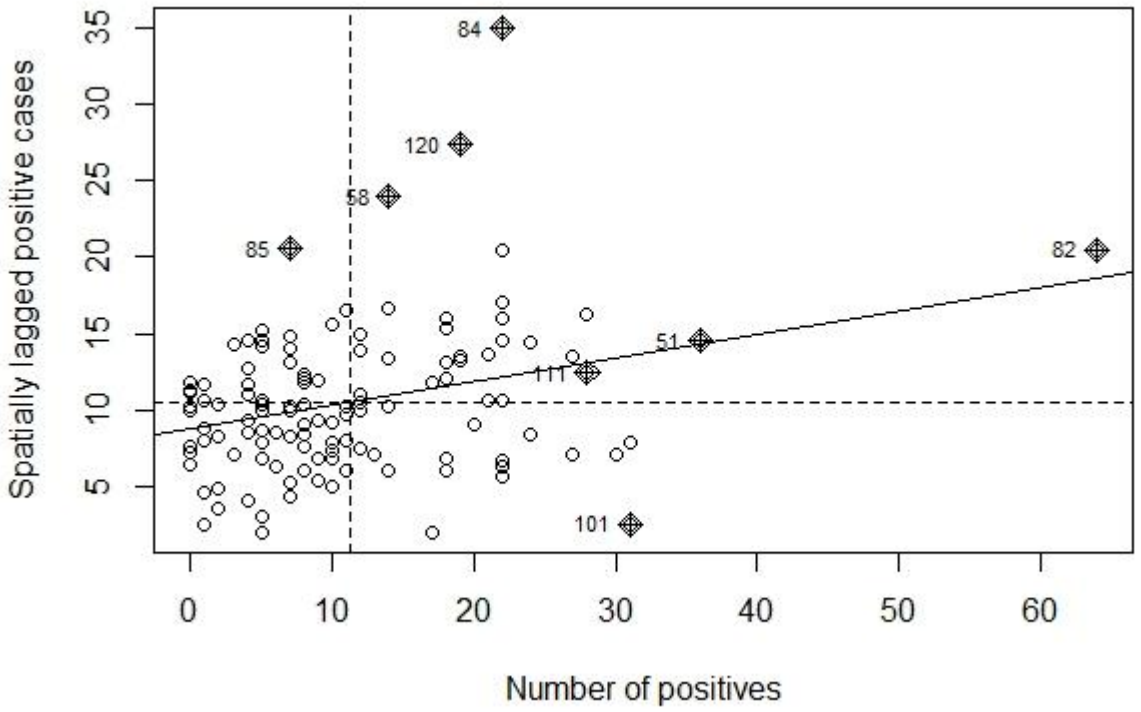
$$I_i = (Y_i - \bar{Y}) \sum_{j=1}^{n} \omega_{ij} (Y_i - \bar{Y}) \qquad (4.2)$$

where *I* is given in equation (4.1).

Table 4.6: Five scenarios of LISA

	High-high	**Hot spots** or locations with high values with similar neighbors
	Low-low	**Cold spots** or locations with low values with similar neighbors
	Low-high	Potential **spatial outliers** or locations with low values with high-values neighbors
	High-low	Potential **spatial outliers** or locations with values with low-value neighbors
	Not significant	Location with **no significant local autocorrelation**

Figure 29: Moran scatterplot for the spatially lagged HIV positive cases per constituency



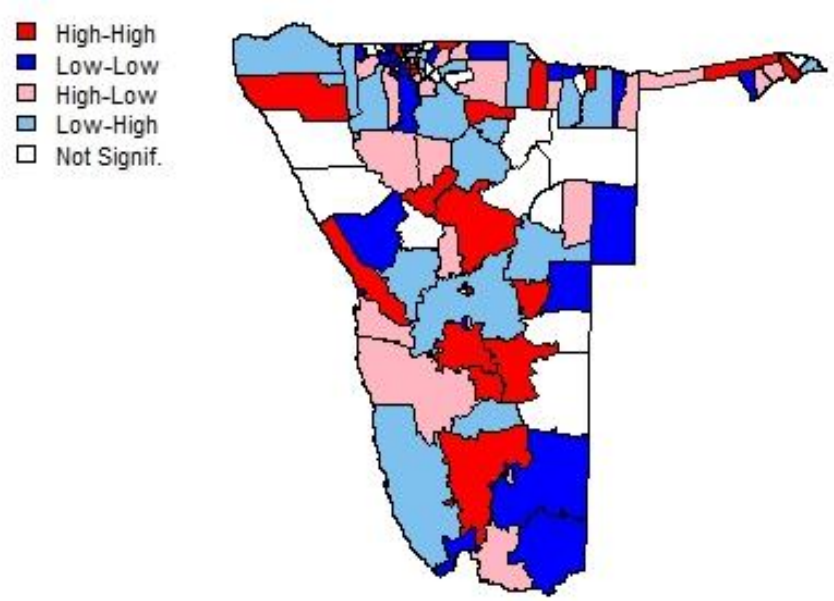
Spatially lagged variables are used with mixed regressive spatially autoregressive estimators. This allows effectively proxy for omitted variables correlated with location. Hence, spatial statistical methods may yield different estimates for various effects than estimators that handle location in a cruder fashion (Elhorst, 2010).

Figure 29 above shows spatially lagged HIV positive cases whereby numbers of positive cases were counted per constituency. Lots of neighbors (number of spatially lagged positive cases) were observed at constituency numbers 84 and 120, while constituency number 101 had few neighbors with high number of HIV positive cases. The local autocorrelation structure may, however, be quite different in some areas from that described by a global statistic such as the Moran's *I*. For example, the constituency number 82 in Figure 29 was characterized by positive spatial autocorrelation with very few neighbors. This happens usually when few neighboring

constituencies are associated with few spatially lagged HIV positive cases while many neighboring constituencies are associated with many spatially lagged HIV positive cases.

The Moran Index is positive when adjacent areas tend to be similar in attributes, negatives when they tend to be more dissimilar than expected, and approximately zero when attribute values are arranged randomly and independently in space (Longley, Goodchild, Maguire, & Rhind, 2015). Moran's *I* conceptual scales of spatial autocorrelation are clustered (for $I > 0$); uncorrelated (for $I = 0$) and contrasting (for $I < 0$).

Figure 30: LISA at constituency level



The Moran's *I* statistic for HIV cases was 0.154 with a standard deviation of 2.807. This shows a significant positive spatial autocorrelation for HIV positive cases with p-value = 0.003, revealing the existence of potential spatial patterns which are clustered in their spatial distributions. For constituencies in Khomas, Hardap, Zambezi, Oshana, Ohangwena and Otjozondjupa regions, pockets of HIV positive cases of high-high spatial clusters were observed (Hot-spots), including a large low-clusters and high-low clusters. On the other hand,

the low-high outliers were mainly located close to the high-high spatial-cluster area (see Figure 30).

It should be noted that the local Moran's Index is sensitive to outliers (Fu, Zhao, Zhang, & Tunney, 2011). Clear clusters were observed mostly from constituencies in Omaheke and Kunene regions. This means that constituencies from these regions have no significant local autocorrelation and this can be attributed to the vastness of these regions and presence of outliers.

4.7. Discussion

This study explored the socio-economic and demographic characteristics as well as sexual behaviors that were associated with HIV prevalence in Namibia. Furthermore, the study applied structured additive regression based on MCMC simulation to Bayesian approach for spatial analysis of HIV prevalence to measure the spatial and random effects of variables that were statistically significant using RR and 95% Crl and were associated with HIV prevalence. Moreover, the measure of spatial similarities between constituencies was quantified using spatial autocorrelation and Moran's *I* statistic was used to measure the relationship between the values of HIV positive cases were based on their locations (constituencies).

Namibia has one of the world's highest HIV prevalence rates. The 2013 Demographic and Health Survey report showed that a national adult population prevalence rate was 14% and as high as 23.7% in Zambezi region. There is a disproportionate distribution of prevalence between women (16.9%) and men (10.9%) aged 15 – 49 years. There were an estimated 217,000 people living with HIV in Namibia of which 166,000 (76%) were receiving

antiretroviral therapy (CDC, 2017). The findings of this study showed that regions in the northern and north-eastern parts of Namibia, namely the Zambezi, Omusati, Oshana, Kavango East, Kavango West and Ohangwena regions, were associated with high HIV prevalence rates. In contrast to this, low prevalence rates were observed in Omaheke, Kunene and Hardap regions of Namibia.

With respect to age, the study found the HIV prevalence to be high in Khomas, Zambezi, Oshana and Omusati region across the youth (those aged 15-35 years). There are many factors that positions this age group to be at high risk of contracting HIV. For example this can be partially explained by poverty, as young people tend to flock to urban areas to look for opportunities to better their livelihoods ending up changing their attitudes and beliefs towards sex when life gets tough (that is, turning into prostitution). According to the 2011 Namibia population and housing census report, highly populated towns in Namibia were Windhoek, Oshakati, Walvis Bay, Katima Mulilo and Outapi, and these towns were located in Khomas, Oshana, Erongo, Zambezi, and Omusati regions respectively. The study variables such as of non-use of condom use during last sexual intercourse, wealth index (poor, middle and richer), living with partner and gender (male) were generally consistent at regional and constituency levels by being significantly associated with HIV prevalence. Various studies pointed out that poverty, social inequality, low income, lack of condom use, no formal education were all associated to high HIV prevalence rates (Harris, 2017; Hollmann, 2016; Musenge et al., 2013; Mwahi, 2014; Ntirampeba et al., 2017; L. Otwombe, 2014) and poor access to health care and TB and HIV control services could also contribute to a high rate of the disease since infectious cases may remain undiagnosed and may not acquire treatment, which could consequently contribute to the transmission dynamics of the disease (Shipanga, 2019)

Spatial effects were obtained both at regional and constituency levels. The models were compared using DIC and the results indicated that random effects (structured spatial effects)

outperformed other models although there was not much difference at constituency level. A significant positive association of covariates to HIV prevalence was observed for condom use, age, marital status, religion, wealth index, level of education and gender in the fixed effect results of Models 1 and 8 respectively. Moreover, the findings of this study confirmed that the posterior mean have a residual effect on HIV prevalence with neighborhood (regions) in white colour showing significant positive effects, black colour showing negative effects and light dark/grey showing non-significant structured and unstructured spatial effects (see Figures 26, 27 and 28).

Finally, the Morans'*I* statistic for HIV cases was 0.15414 with a standard deviation of 2.807 and a p-value of 0.003. The results proved that Morans'*I* test was significant with positive spatial autocorrelation for HIV cases. The LISA showed hot spots for some constituencies in Erongo, Khomas, Omusati, Oshana and Ohangwena regions while locations with no significant local autocorrelation were observed in mostly parts of Omaheke and Kunene regions. Some constituencies had few neighbors with high number of HIV positive cases and this complicates the interpretation (for example constituency number 84 in Figure 29) because HIV positive cases co-varies with constituencies, which were difficult to quantify sometimes due to other omitted variables correlated with location. Fitting the Moran's *I* statistic to the data was more important because the results gave a clear picture of which constituencies were spatially correlated and clustered given the HIV positive cases. In summary, the autocorrelation of HIV cases was clustered in some constituencies of Erongo, Khomas, Omusati, Oshana and Ohangwena regions.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

This study explored the importance of semi-parametric Bayesian approach to model the factors associated with HIV prevalence in Namibia.

HIV prevalence for the respondents aged 15 - 24 years is usually assumed to represent new infections and therefore serves as a proxy for HIV incidence. The findings of this study showed that the type of residence, lack of condom use, level of education, sex and age were associated with high rate of HIV prevalence in most areas of the northern and northeastern regions of Namibia, namely the Zambezi, Kavango East, Kavango West, Ohangwena, Omusati and Oshana regions. The RR of the coefficients of non-condom use, gender (male), marital status (living with partner), level of education (primary education), Wealth index (poor, middle, richer) and religion (ELCIN), and 95% CrI are highly significant in the Bayesian context and, thus positively associated with HIV prevalence in Namibia.

There was a positive influence of fixed effects on HIV prevalence in the northern part of Namibia with positive structured and unstructured spatial effects specifically in Zambezi, Kavango East, Kavango West, Omusati and Oshana. Erongo and Khomas regions also had pockets of constituencies with positive spatial random effects, while a negative influence on HIV prevalence was observed more in Omaheke and Hardap regions. Similar results were obtained by Ntirampeba et al. (2017), whose approach involved modelling different spatial levels in a single model by joining two sources of data. However, the results by Ntirampeba et al. (2017) were based on a bivariate modelling approach using the spatial shared component model that jointly fitted information from the NHSS data to the NDHS data when they predicted the HIV prevalence in Namibia. This study used a Poisson modelling approach and

identified risk factors associated with HIV prevalence using one data source and the results showed the difference of variations between spatially structured and unstructured random effects.

DIC was used as a measure to select the best fit model and the results showed that the models with a spatial structure effect were the best fit at regional and constituency levels and were better at explaining the HIV prevalence rate. The unstructured spatial effects models were close to best fit model with minimal deviance. Thus, there was evidence that neighbouring areas have different variations of HIV prevalence rate.

The results of Moran's *I* statistic showed that there was significant spatial clustering (Khomas, Hardap, Zambezi, Oshana, Ohangwena and Otjozondjupa regions) of HIV positive cases in Namibia. Although hot spots of HIV positive cases were identified, it must be noted that spatial autocorrelation is subject to outliers. Understanding socio-economic demographic characteristics and sexual behavior that were associated with HIV/AIDS prevalence using spatial analysis may contribute further to broad knowledge of decision makers to enable them to establish better detection of HIV.

5.2. Recommendations

Hot spots (clustering) or locations with high values of HIV positive cases with similar neighbors were identified as Erongo, Khomas, Omusati, Oshana and Ohangwena region (Figure 30). Therefore, future research should focus more on these areas when identifying risk factors associated to HIV prevalence in Namibia using spatial modelling approaches. The study also recommends future research to use census data when modelling HIV prevalence in Namibia using Bayesian multi-scale modelling of convolution model (BYM) that models the relative risks as a function of spatially structured and unstructured random effects.

NSA is tasked to take the lead in formalising the National Spatial Data Infrastructure (NSDI) for Namibia. Quality, timely and accurate spatial data production on diseases is still a challenge and it will be helpful if NSA starts to intergrate the available diseases data to the NSDI to promote efficient production, use, maintenance and dissemination of relevant, quality and accurate spatial information that is fit-for-purpose, particularly in providing evidence-based decision making at all levels of society. This can help in achieving the desired outcome of the Health and Nutrition strategies for the NDP5 in accelerating health infrastructure development and resource management which is intended to result in quality health care provision to all Namibians.

The demographics, socio-economic and cultural factors for HIV prevalence in Namibia should be investigated further at multiscale level to improve on the analysis and reporting methods because the NDHS data does not cover the entire population and this affect the process of spatial analysis. Using census data could reveal most important aspects of spatial structures of HIV prevalence in Namibia that would be exceedingly critical for both illustrative as well as policy implementation purposes.

Since funding of HIV/AIDS programmes in Namibia is subject to numerous competing national priorities, utilizing the existing financial resources requires a better understanding of the heterogeneity, transmission dynamics, and geographic variation of epidemics in the country, together with the use of evidence based interventions.

REFERENCES

Annis, J., Miller, B. J., & Palmeri, T. J. (2017). Bayesian inference with Stan: A tutorial on adding custom distributions. *Behavior research methods, 49*(3), 863-886. doi:10.3758/s13428-016-0746-9

Aregay, M., Lawson, A. B., Faes, C., & Kirby, R. S. (2017). Bayesian multi-scale modeling for aggregated disease mapping data. *Statistical methods in medical research, 26*(6), 2726-2742.

Atilola, G. (2014). *Prevalence, spatial patterns and factors associated with HIV infection in Zimbabwe, 2011*.

Belitz, C., Brezger, A., Kneib, T., Lang, S., & Umlauf, N. (2009). BayesX-Software for Bayesian inference in structured additive regression models. *URL http://www. stat. uni-muenchen. de/~ bayesx*.

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.

Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*: John Wiley & Sons.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and spatio-temporal epidemiology, 4*, 33-49.

CDC. (2017). Overview of CDC in Namibia.

Chimoyi, L. A., & Musenge, E. (2014). Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011. *BMC public health, 14*(1), 555.

Congdon, P. (2014). *Applied bayesian modelling* (Vol. 595): John Wiley & Sons.

Cuadros, D. F., Li, J., Branscum, A. J., Akullian, A., Jia, P., Mziray, E. N., & Tanser, F. (2017). Mapping the spatial variability of HIV infection in Sub-Saharan Africa:

Effective information for localized HIV prevention and control. *Scientific Reports, 7*(1), 9093. doi:10.1038/s41598-017-09464-y

De la Torre, C., Khan, S., Eckert, E., Luna, J., & Koppenhaver, T. (2009). HIV/AIDs in Namibia: Behavioral and contextual factors driving the epidemic.

Denis, P. (2013). HIV/AIDS and religion in sub-Saharan Africa: an emerging field of enquiry. *Archives de sciences sociales des religions*(164), 43-58.

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. *Spatial economic analysis, 5*(1), 9-28.

Fu, W., Zhao, K., Zhang, C., & Tunney, H. (2011). Using Moran's I and geostatistics to identify spatial patterns of soil nutrients in two different long-term phosphorus-application plots. *Journal of Plant Nutrition and Soil Science, 174*(5), 785-798. doi:10.1002/jpln.201000422

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The incorporated statistician, 5*(3), 115-146.

Gelaw, Y. A., Magalhães, R. J. S., Assefa, Y., & Williams, G. (2019). Spatial clustering and socio-demographic determinants of HIV infection in Ethiopia, 2015–2017. *International Journal of Infectious Diseases, 82*, 33-39.

Habyarimana, F., Zewotir, T., Ramroop, S., & Ayele, D. (2016). Spatial Distribution of Determinants of Malnutrition of Children under Five Years in Rwanda: SimultaneousMeasurement of Three Anthropometric Indices. *Journal of Human Ecology, 54*(3), 138-149.

Harris, T. R. (2017). *Multiscale spatial modelling of diabetes and hypertension in Namibia* (Master of Science in Applied Statistics & Demography), University of Namibia, Unam respiratory. Retrieved from http://www.unam.edu.na/library

Hespanhol, L., Vallio, C. S., Costa, L. M., & Saragiotto, B. T. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian journal of physical therapy, 23*(4), 290-301.

Hollmann, L. (2016). Mapping HIV in Uganda.

International, T. N. M. o. H. a. S. S. M. a. I. (2014). *The Namibian Demographic and Health Survey 2013*. Retrieved from Windhoek, Namibia, and Rockville, Maryland.:

Khana, D., Rossen, L. M., Hedegaard, H., & Warner, M. (2018). A Bayesian Spatial and Temporal Modeling Approach to Mapping Geographic Variation in Mortality Rates for Subnational Areas With R-Inla *Journal of data science : JDS, 16*(1), 147-182.

Kilmarx, P. H. (2009). Global epidemiology of HIV. *Current Opinion in HIV and AIDS, 4*(4), 240-246.

Lawson, A. B. (2013a). *Bayesian Disease Mapping Hierarchical Modeling in Spatial Epidemiology*. LLC: Taylor & Francis Group.

Lawson, A. B. (2013b). *Statistical methods in spatial epidemiology*: John Wiley & Sons.

Liu, C., & Sharma, A. (2018). Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Analytic methods in accident research, 17*, 14-31.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic information science and systems*: John Wiley & Sons.

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

Mabaso, M., Sokhela, Z., Mohlabane, N., Chibi, B., Zuma, K., & Simbayi, L. (2018). Determinants of HIV infection among adolescent girls and young women aged 15–24 years in South Africa: a 2012 population-based national household survey. *BMC public health, 18*(1), 183.

Menges, W. (2018, 2018-02-12). Global Fund cuts deepen Namibia's woes, National. *The Namibian*, p. 1. Retrieved from https://www.namibian.com.na/174350/archive-read/Global-Fund-cuts-deepen-Namibias-woes

MoHSS. (2016). *Surveillance Report of the 2016 National HIV Sentinel Survey*. Retrieved from www.mhss.gov.na

MoHSS, & ICF-International. (2014). *The Namibia demographic and health survey 2013*. Retrieved from Windhoek, Namibia, and Rockville, Maryland, USA:

Moran, P. (1947). *Random associations on a lattice.* Paper presented at the Mathematical Proceedings of the Cambridge Philosophical Society.

Musenge, E., Vounatsou, P., Collinson, M., Tollman, S., & Kahn, K. (2013). The contribution of spatial analysis to understanding HIV/TB mortality in children: a structural equation modelling approach. *Global Health Action, 6*, 10.3402/gha.v3406i3400.19266. doi:10.3402/gha.v6i0.19266

Mwahi, E. M. (2014). *Shared-component model with application to mapping gender specific pattern in HIV testing and condom use in Namibia.*

NSA. (2011). *Namibia 2011: Population & Housing Census Main Report*. Retrieved from

Ntirampeba, D. (2018). *Modelling Spatio-temporal patterns of disease risk for data with misalignment and measurement errors: An application on measles and HIV prevalence data in Namibia.* (Doctor of Philosophy in Science (Statistics)), University of Namibia, Unam respiratory. Retrieved from http://www.unam.edu.na/library

Ntirampeba, D., Neema, I., & Kazembe, L. N. (2017). Joint spatial modelling of disease risk using multiple sources: an application on HIV prevalence from antenatal sentinel and demographic and health surveys in Namibia. *Global Health Research and Policy, 2*, 22. doi:10.1186/s41256-017-0041-z

Okango, E., Mwambi, H., & Ngesa, O. (2016). Spatial modeling of HIV and HSV-2 among women in Kenya with spatially varying coefficients. *BMC Public Health, 16*(1), 355. doi:10.1186/s12889-016-3022-0

Otwombe, L. (2014). *Spatial distribution and analysis of factors associated wiyh HIV infection among young people in Eastern Africa: applied to the MEASURE demographic and health survey data collected between 2007 and 2011*.

Otwombe, L. A. (2013). Spatial distribution and analysis of factors associated wiyh HIV infection among young people in Eatern Africa: Applied to the measure demographic and health survey data collected between 2007 and 201.

Paradis, E. (2018a). Moran's autocorrelation coefficient in comparative methods. *Consultado el, 7*.

Paradis, E. (2018b). Moran's autocorrelation coefficient in comparative methods.

Pereira, S., Turkman, F., & Correia, L. (2017). Spatio-temporal analysis of regional unemployment rates: A comparison of model based approaches. *arXiv preprint arXiv:1704.05767*.

Plant, R. E. (2012). *Spatial data analysis in ecology and agriculture using R*. University of Califorinia: CRC Press, Taylor & Francis Group.

Poh-Chin, L., Fun-Mun, S., & Ka-Wing, C. (Eds.). (2009). *Spatial Epidemiological Approaches in Disease Mapping and Analysis*. Boca Raton London New York: Taylor & Francis Group.

Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research, 25*(4), 1145-1165.

Rotheram-Borus, M. J., Davis, E., & Rezai, R. (2018). Stopping the rise of HIV among adolescents globally. *Current opinion in pediatrics, 30*(1), 131-136.

Schrödle, B., & Held, L. (2011). Spatio-temporal disease mapping using INLA. *Environmetrics, 22*(6), 725-734.

Services, M. o. H. a. S. (2014). Namibia Demographic and Health Survey 2013. *Namibia Demographic and Health Survey 2013.*

Shipanga, A. I. (2019). *A spatial temporal analysis of survival among TB and HIV co-infected patients in Erongo region, Namibia.* University of Namibia.

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.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(3), 485-493.

Srinivasan, R., & Venkatesan, P. (2015). Bayesian spatio-temporal model for tuberculosis in India. *The Indian journal of medical research, 141*(4), 478.

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

UNAIDS. (2016). Report on the Global HIV/AIDS Epidemic.

UNAIDS. (2018, 31 July 2019). The Global HIV/AIDS Epidemic. Retrieved from https://www.unaids.org/en/resources/fact-sheet

Venkatesan, P., & Srinivasan, R. (2008). Bayesian model of HIV/AIDS in India: A spatial analysis. *Applied Bayesian Statistical Analysis*.

Wang, X., Ryan, Y. Y., & Faraway, J. J. (2018). *Bayesian Regression Modeling with INLA*: Chapman and Hall/CRC.

WBG. (2018a). Country profile: Namibia. Retrieved from https://data.worldbank.org/country/namibia

WBG. (2018b, Oct 04, 2019). The World Bank In Namibia. Retrieved from https://www.worldbank.org/en/country/namibia/overview

WHO. (2018). *Report on the HIV/AIDS Epidemic in Sub Saharan Africa*. Retrieved from

Wilson, D., & Taaffe, J. (2017). Tailoring the local hiV/aiDs response to local hiV/aiDs epidemics (Vol. 6): World Bank Washington, DC.

Woldemariame, S. H. (2013). *Factors Determining the Prevalence of HIV/AIDS in Ethiopia.* Master's Thesis in Mathematical Statistics: Stockholms universitet.

Zvoushe, I. (2014). The determinants of HIV and AIDS among reproductive age group in Zimbabwe: a comparative study between 1999 and 2005 Zimbabwe demographic health surveys.

ANNEX 1: R-Codes

R-codes for study site

```
install.packages(c("dismo","ggplot2","rgbig","ggmap","maptools
","GISTools","mapdata","ggsn"))
library(dismo);library(ggplot2);library(rgbif)
library(ggmap);library(maptools);library(GISTools)
library(mapdata);library(ggsn)
library(foreign);library(sp)

geo<-read.csv("file:///C:/Users/tshinyemba/Desktop/Job Shikong
o/HIV_Coordinates.csv",header=T)
head(geo)
mymap<-qmap(location="Namibia",zoom = 6,maptype="toner",source
= c("stamen"),title="Study Site")
mymap
mymap+geom_jitter(data=geo,aes(geo$LONGNUM,geo$LATNUM),alpha=0
.6, size = 2, color = "red")
north2(mymap+geom_jitter(data=geo,aes(geo$LONGNUM,geo$LATNUM),
alpha=0.6, size = 2, color = "red"),0.06,0.15,scale = 0.2, sym
bol = 1)
```

R - Codes for Moran I

```
setwd ("C:\Users\ShikongoJN\Desktop\Academics")
#load packages
library(rgeos)
library(rgdal)
```

```
library(sp)
library(ggplot2)
library(ggmap)
library(dplyr)
library(raster)
library(leaflet)
library(spatstat)
library(spdep)
library(rgdal)
library(latticeExtra)
library(RColorBrewer)
library(gridExtra)
library(Matrix)
library(lattice)
library(maptools)
library(foreign)
library(ggmap)
library(BayesX)
library(spData)
library(INLA)
```

Load in shapefile

```
Namibia <- readOGR(dsn = ".", layer= "ADMIN_Constituency_Bound
aries_2014")
plot(International)
DHS <- read.csv(file="CB.csv", header=TRUE, sep=",")
edit(DHS)
queen.nb = poly2nb(International)
summary(queen.nb)
```

```
queen.listw=nb2listw(queen.nb) #convert nb to listw type
listw=queen.listw
# plot neighbourhood
plot(Namibia, border=gray(.5))
plot(queen.nb, coordinates(Namibia), add=TRUE)
# Merge the two data files
data2 <- merge(Namibia, DHS, by='CONST')
# Computing moran's I in R for spatial Data
mi <- moran.test(data2$Positives, listw = nb2listw(queen.nb))
mi

# Plot Moran I scatter plot, Moran I (local),and probability o
f most significant Moran I (Chapter 4, Figure )
moran.plot(data2$Positives, listw = queen.listw, xlab="Number o
f positives", ylab="Spatially lagged positive cases")
locm <- localmoran(data2$Positives, listw = nb2listw(queen.nb)
)
summary(locm)
data2$Positives <- scale(data2$Positives)
data2$Positives <- lag.listw(listw, data2$Positives)
summary(data2$Positives)
summary(data2$lag_positives)
data2$quad_sig <- NA
data2@data[(data2$Positives >= 0 & data2$lag_Positives >= 0) &
(locm[, 5] <= 0.05), "quad_sig"] <- 1
data2@data[(data2$Positives <= 0 & data2$lag_Positives <= 0) &
(locm[, 5] <= 0.05), "quad_sig"] <- 2
data2@data[(data2$Positives >= 0 & data2$lag_Positives <= 0) &
(locm[, 5] <= 0.05), "quad_sig"] <- 3
```

```
data2@data[(data2$Positives >= 0 & data2$lag_Positives <= 0) &
(locm[, 5] <= 0.05), "quad_sig"] <- 4
data2@data[(data2$Positives <= 0 & data2$lag_Positives >= 0) &
(locm[, 5] <= 0.05), "quad_sig"] <- 5
breaks <- seq(1, 5, 1)
labels <- c("High-High", "Low-Low", "High-Low", "Low-High", "N
ot Signif.")
np <- findInterval(data2$quad_sig, breaks)
colors <- c("red", "blue", "lightpink", "skyblue2", "white")
plot(data2, col = colors)
legend("topleft", legend = labels, fill = colors, bty = "n", c
ex =0.7)
plot(variogram(data2$Positives ~ 1, locations = coordinates(da
ta2), data = data2),
     type = "b", pch = 16, main = "Variogram of Constituency")
```

ANNEX 2: BayesX CODES

Codes for structured and unstructured spatial effects of HIV prevalence

```
> defaultpath = c:\data2020
%Dataseb object
> dataset hivreduced
%Read in dataset into dataset object in object called tbhiv
> hivreduced.infile using C:\data2020\reduced2020.txt
%Describe the dataset and view the dataset
> hivreduced.describe
%loading in the map object
> map h
h.infile using C:\data2020\NA.csv
> h.reorder

> map h1
> h1.infile using C:\data2020\namibia.csv
> h1.reorder
%Plot the map object and view the map
> h.describe
> h1.describe
> h.outfile, replace graph using c:\data2020\namibiasort.bnd
> h1.outfile, replace graph using c:\data2020\namibiaconsort.bnd

> bayesreg m1
```

> m1.outfile=c:\data2020\m1

> m1.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural, iterations=12000 burnin=2000 step=10 family= poisson predict using hivreduced

> bayesreg m2

> m2.outfile=c:\data2020\m2

> m2.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural + Region(spatial,map=h), iterations=12000 burnin=2000 step=10 family=poisson predict using hivreduced

> bayesreg m3

> m3.outfile=c:\data2020\m3

> m3.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural + Region(random), iterations=12000 burnin=2000 step=10 family=poisson predict using hivreduced

> bayesreg m4

> m4.outfile=c:\data2020\m4

> m4.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural + Region(spatial,map=h) + Region(random), iterations=12000 burnin=2000 step=10 family=poisson predict using hivreduced

> bayesreg m5

> m5.outfile=c:\data2020\m5

> m5.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural + ID(spatial,map=h1), iterations=12000 burnin=2000 step=10 family=poisson predict using hivreduced

> bayesreg m6

> m6.outfile=c:\data2020\m6

> m6.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural + ID(random), iterations=12000 burnin=2000 step=10 family=poisson predict using hivreduced

> bayesreg m7

> m7.outfile=c:\data2020\m7

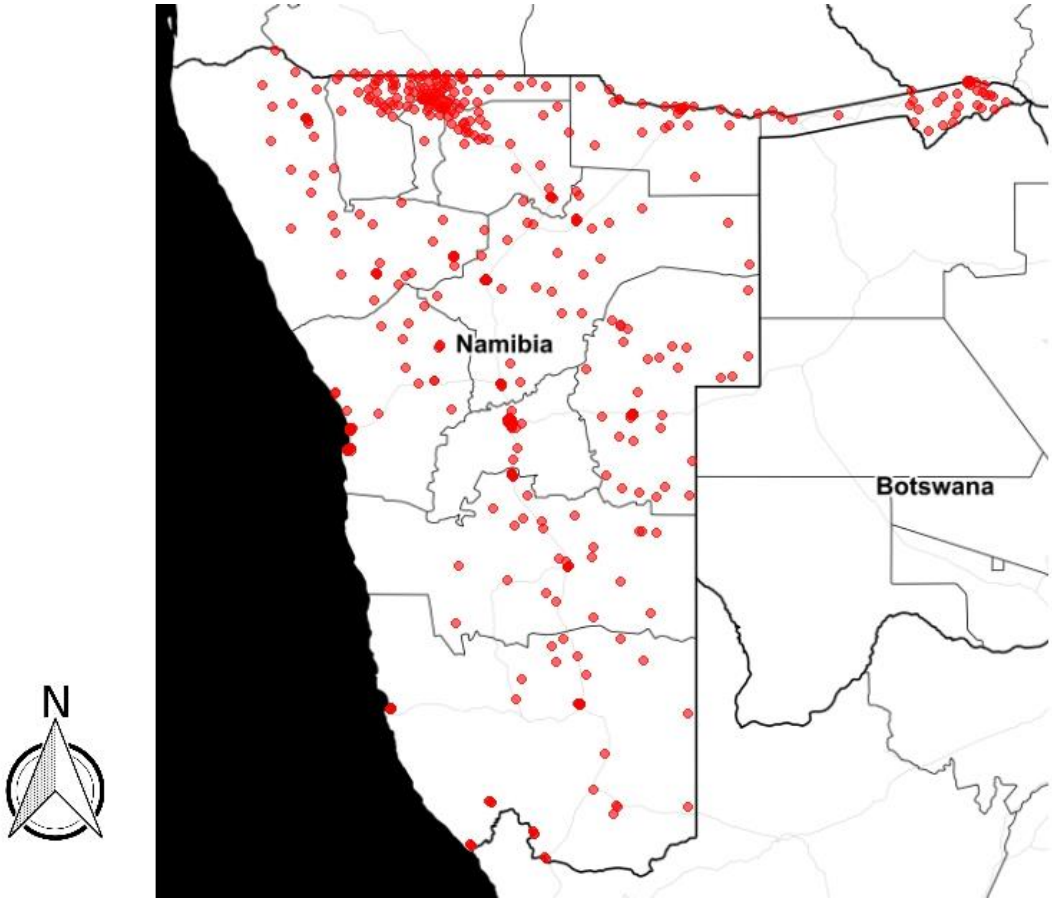
> m7.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural + ID(spatial,map=h1) + ID(random), iterations=12000 burnin=2000 step=10 family=poisson predict using hivreduced

> bayesreg m8

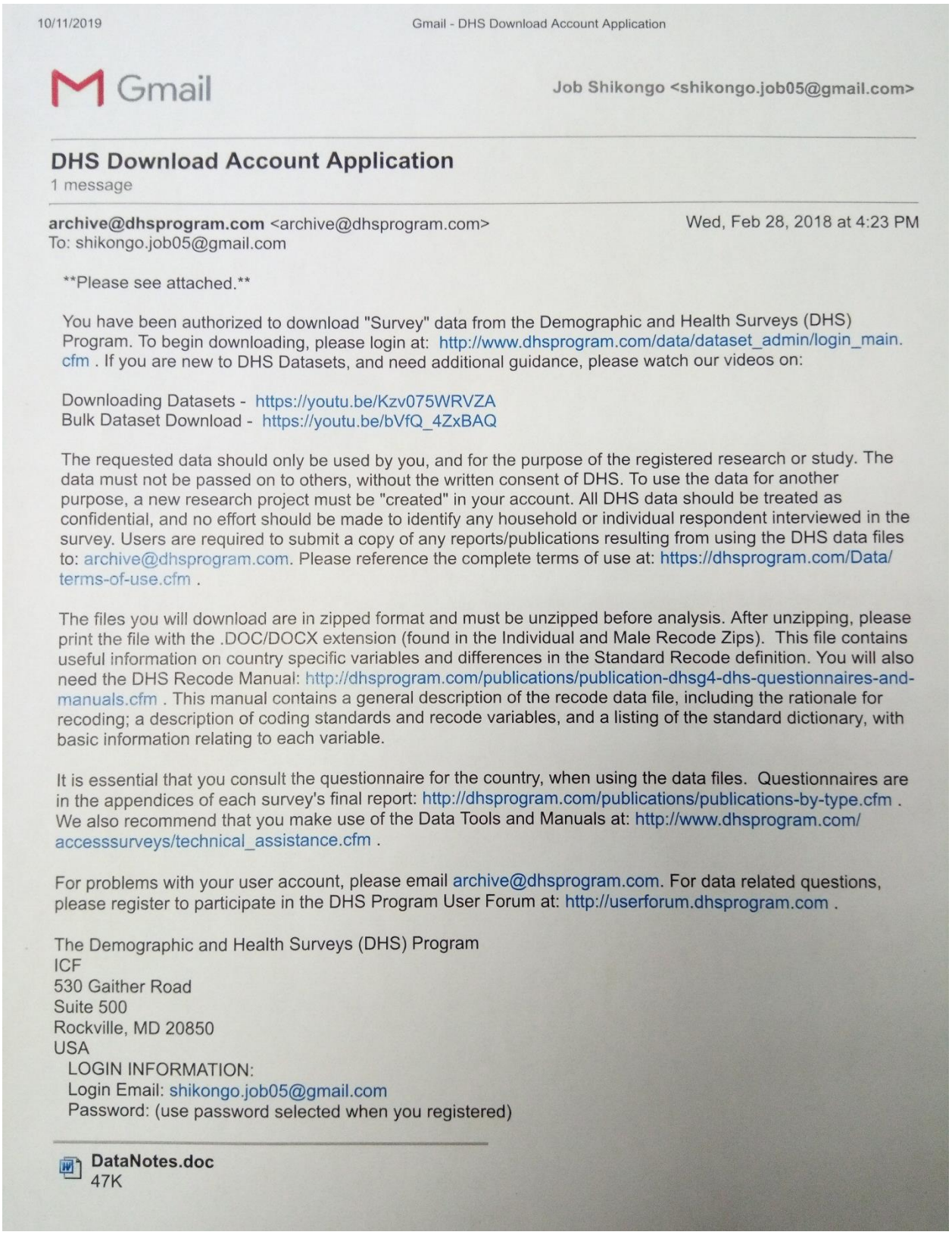
> m8.outfile=c:\data2020\m8

> m8.regress HIV = age + gendermale + mstatus1 + mstatus2 + mstatus3 + mstatus4 + mstatus5 + edu1 + edu2 + edu3 + WealthIndexP + WealthIndexM + WealthIndexR + religion1 + religion2 + religion3 + religion4 + condomUse + residtyperural + Region(spatial,map=h) + Region(random) + ID(spatial,map=h1) + ID(random), iterations=12000 burnin=2000 step=10 family=poisson predict using hivreduced

ANNEX 3: Study site HIV testing



ANNEX 4: Authorization to download DHS data



10/11/2019 Gmail - DHS Download Account Application

Gmail

Job Shikongo <shikongo.job05@gmail.com>

DHS Download Account Application

1 message

archive@dhsprogram.com <archive@dhsprogram.com> Wed, Feb 28, 2018 at 4:23 PM
To: shikongo.job05@gmail.com

Please see attached.

You have been authorized to download "Survey" data from the Demographic and Health Surveys (DHS) Program. To begin downloading, please login at: http://www.dhsprogram.com/data/dataset_admin/login_main.cfm . If you are new to DHS Datasets, and need additional guidance, please watch our videos on:

Downloading Datasets - https://youtu.be/Kzv075WRVZA
Bulk Dataset Download - https://youtu.be/bVfQ_4ZxBAQ

The requested data should only be used by you, and for the purpose of the registered research or study. The data must not be passed on to others, without the written consent of DHS. To use the data for another purpose, a new research project must be "created" in your account. All DHS data should be treated as confidential, and no effort should be made to identify any household or individual respondent interviewed in the survey. Users are required to submit a copy of any reports/publications resulting from using the DHS data files to: archive@dhsprogram.com. Please reference the complete terms of use at: https://dhsprogram.com/Data/terms-of-use.cfm .

The files you will download are in zipped format and must be unzipped before analysis. After unzipping, please print the file with the .DOC/DOCX extension (found in the Individual and Male Recode Zips). This file contains useful information on country specific variables and differences in the Standard Recode definition. You will also need the DHS Recode Manual: http://dhsprogram.com/publications/publication-dhsg4-dhs-questionnaires-and-manuals.cfm . This manual contains a general description of the recode data file, including the rationale for recoding; a description of coding standards and recode variables, and a listing of the standard dictionary, with basic information relating to each variable.

It is essential that you consult the questionnaire for the country, when using the data files. Questionnaires are in the appendices of each survey's final report: http://dhsprogram.com/publications/publications-by-type.cfm . We also recommend that you make use of the Data Tools and Manuals at: http://www.dhsprogram.com/accesssurveys/technical_assistance.cfm .

For problems with your user account, please email archive@dhsprogram.com. For data related questions, please register to participate in the DHS Program User Forum at: http://userforum.dhsprogram.com .

The Demographic and Health Surveys (DHS) Program
ICF
530 Gaither Road
Suite 500
Rockville, MD 20850
USA
LOGIN INFORMATION:
Login Email: shikongo.job05@gmail.com
Password: (use password selected when you registered)

DataNotes.doc
47K