Spatial Distribution of HIV/AIDS in Botswana

T. Mupundu

Thesis submitted for the degree *Philosophiae Doctor* in Statistics at the North-West University

Promoter: Prof N.D. Moroke
Co-promoter: Dr. F. Matarise

Graduation October 2017
Student number: 25823310
http://dspace.nwu.ac.za/

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

It all starts here ™

Acknowledgements

First and foremost I would like to express my gratitude to my supervisors Professor Moroke and Dr. Matarise for their guidance and support. They have guided and tremendously assisted me during the writing of this report.

My appreciation goes to my husband and my children, Tendai, Tinashe, Thokozani and my granddaughter Christine for encouraging and supporting me. Thank you to my mother, brothers and sister for your encouragement and support.

Various people have played significant roles in my studies, without their support I would not have been able to do this work. I would like to acknowledge and extend my gratitude to the following people:

- Prof. Nehemia Mavetera for continuously encouraging and assisting me.
- Mr. G. Mlambo for assisting me in learning GIS and for encouraging and supporting me throughout my studies.
- Mr. G Koontse, thank you for assisting me in acquiring the GIS software.
- I would like to thank the Central Statistics of Botswana for providing and allowing me to use 2013 BAIS IV data for my study.
- I will not fail to thank my friend Claris, who has been my pillar from the day we registered. Thank you my friend for your support academically and spiritually. May God bless you.
- To all my friends too numerous to mention both in Zimbabwe and here in Botswana, I say a big thank you for your good wishes and prayers. May God bless you all.

Table of contents

ABSTRACT x
STUDY ORIENTATION 1
1.1. Introduction 1
1.2. Background of the study 2
1.3. Importance of Spatial Analysis 3
1.4. Problem Statement 4
1.5. Rationale of the study 5
1.6. Aim of study 6
1.7. Research Questions 6
1.8. Research Objective(s) 6
1.9. Significance of the Study 6
1.10. Assumptions 8
1.11. Outline of the thesis 8
1.12. Definition of terms 9
1.13. Conclusion 9
CHAPTER 2 10
LITERATURE REVIEW 10
2.1. Introduction 10
2.2. HIV/AIDS Worldwide 10
2.2.1. Modes of transmission 14
2.2.2. Sexual transmission 14
2.2.3. Parental transmission 15
2.2.4. Mother-to-child transmission 15
2.3. HIV/AIDS in Africa 16
2 3.1. Impact of HIV/AIDS in Africa 20
2.4. Factors affecting the spread 22
2.5. HIV/AIDS in Botswana 23
2.5.1. Challenges 25
2.5.2. Key affected populations in Botswana 26

2.5.3. Determinants of HIV/AIDS in Botswana 26
2.6. Use of GIS in Health Research 28
2.7. Limitations of Geographic Information System 32
2.7.1 Problems to do with the GIS data model 32
2.7.2 Problems to do with data itself 32
2.7.3 Problems with academic paradigm 33
2.7.4 Practical problems. 33
2.8. Spatial Analysis 33
2.8.1. Types of spatial statistics 34
2.9. Literature on Spatial Analysis 34
2.10. Spatial-Temporal analysis 37
2.11. Interpolation Methods 41
2.12. Conclusion 43
CHAPTER 3 45
RESEARCH DESIGN AND METHODOLOGY 45
3.1. Introduction 45
3.2. Research design 45
3.3. Nature of methodology 46
3.4. Philosophical Perspective 47
3.5. Philosophical Approaches in Social Sciences 48
3.5.1. Positivism 48
3.5.2. Social constructionism 49
3.5.3. Critical paradigm 49
3.5.4. Postmodernism 49
3.5.5. Deductive approaches 50
3.5.6. Quantitative approaches 51
3.6. Study Area 52
3.7. Data 53
3.8. Data Collection 54
3.9. Multi- stage sampling design 55
3.10. Conceptual Framework 57

3.11. Methods of data analysis 58
3.10.1 Preliminary data analysis 59
3.10.2 Descriptive statistics 59
3.10.3 Test for reliability 59
3.10.4 Primary data analysis 60
3.10.5 Spatial Autocorrelation of HIV/AIDS prevalence in Botswana 61
3.10.6 Getis-Ord General G statistic 62
3.10.7 Identification of Hotspot 63
3.10.8 Getis-OrdGi*statistics 64
3.10.9 Kulldorff's spatial scan statistics 66
3.10.10 Spatial Interpolation methods 68
3.10.11 Inverse distance interpolation Method 69
3.10.12 Kriging Interpolation Method 71
3.10.13 Variogram 73
3.10.14 Natural Neighbour Interpolation Method 74
3.12. Methods of comparing Interpolation techniques 76
3.13. Logistic Regression 77
3.13.1 Generalised Linear Model 78
3.13.2 Evaluating goodness of fit 79
3.14. Ethical consideration 81
3.15. Conclusion 81
CHAPTER 4 **82**
DATA ANALYSIS AND RESULTS **82**
4.1. Introduction 82
4.2. Descriptive Statistics 82
4.3. Spatial analysis 91
4.3.1. Exploratory spatial data analysis 92
4.3.2. Autocorrelation analysis 92
4.3.3. Getis-Ord general G 93
4.3.4. Hot Spot Analysis results 93
4.3.5. Local Indicator of Spatial Association results 94

Table of Tables

Table of contents ii

Table 3.1 Philosophical Approaches 50

Table 3.2 Deductive Research 51

Table 4.1 Distribution of HIV positive participants per district 83

Table 4.2 HIV/AIDS positive individuals by Gender 84

Table 4.3 Distribution of HIV/AIDS individuals by Marital status 85

Table 4.4 Distribution of HIV/AIDS individuals by Education 87

Table 4.5 Distribution of HIV/AIDS positive individuals by Religion 88

Table 4.6 Distribution of HIV/AIDS positive individuals by Employment. 90

Table 4 7. Moran's Index for spatial autocorrelation of HIV/AIDS 93

Table 4.8 Getis-Ord General G 93

Table 4.9 HIV/AIDS clusters in Botswana districts using Kulldorff Sat scan 95

Table 4.10 Prediction errors for different kriging methods 98

Table 4.11 Showing RMSE for the three interpolation methods 100

Table 4.12 Risk factors associated with HIV/AIDS in Gaborone district 102

Table 4.13 HIV/AIDS factors in Francistown 103

Table 4.14 Risk factors associated with HIV/AIDS prevalence in Selebi-Phikwe, Central Bobonong, Central Boteti, Kagtleng, Southeast and Ngamiland South 104

Table 4.15 Risk factors associated with HIV/AIDS prevalence in Kweneng District

Table 4.16 Risk factors associated with HIV/AIDS in Central Serowe and Central Tutume 107

Table 4.17 Risk factors associated with HIV/AIDS Central Mahalapye district 108

Table 4.18 -2Loglikelihood ratio test and R^2 109

Table of Figures

Figure 1.1 HIV/AIDS Prevalence Rate by Age and Gender (Adapted from BAIS IV) 2

Figure 2.1 Adult HIV/AIDS Prevalence Rate for the Year 2014 (Source: Kaiser Family Foundation, 2015).. 12

Figure 2.2 Total number of people living with HIV/AIDS for the year 2014 (Source: Kaiser Family Foundation, 2015) 13

Figure 2.3 Number of new infection for the year 2014 (Source: Kaiser Family......... 14

Foundation, 2015) 14

Figure 2.4 Prevention progress of PMTCT in selected priority countries in Africa. Source: UNAIDS 2016. On the fast track to an AIDS free generation 16

Figure 2.5 Number of people living with HIV/AIDS for the year 2014 (Source: Kaiser Family Foundation, 2015).. 17

Figure 2.6 Adult Prevalence Rate for the year 2014 (Source: Kaiser Family Foundation, 2015) 18

Figure 2.7 Number of New Infections for the year 2014 (Source: Kaiser Family Foundation, 2015) 18

Figure 2.8 Number of Deaths for the year 2014 (Source: Kaiser Family Foundation, 2015) 19

Figure 2.9 Statistics and facts about HIV in Africa (Adapted from: AIDS in Africa, 2013. Medwiser... 20

Figure 3.1 Research philosophy in the 'research onion' Source: Saunders *et al.* (2009) 46

Figure 3.2: The interrelationship between ontology, epistemology, methodology, methods and data Source: Grix (2002:180) 48

Figure 3.3 Botswana's Political map. Source: www.mapsofworld.com-2013 52

Figure 3.4 Shape file map showing 28 census districts of Botswana: Source: Statistics Botswana 54

Figure 3.5 Conceptual frameworks for factors contributing to the spread of HIV/AIDS 58

Figure 3.6 Framework for the guide of spatial analysis. Adopted from Geospatial analysis of global health, 2015.. 60

Figure3.7 Natural Neighbour interpolation of a point 76
Figure 4.1 showing the spatial distribution of HIV/AIDS in Botswana 83
Figure 4.2Distribution of HIV/AIDS in selected Botswana districts by Age 86
Figure 4.3 Histogram showing distribution of HIV/AIDS 92
Figure 4.4 LISA Cluster Map and LISA Significance Map (For district names refer to Figure 3.4) 94
Figure 4.5 Gi* Cluster Map and Significant Map (For district names refer to Figure 3.4) 96
Figure 4.6 Maps of Ordinary, Simple and Universal kriging for HIV/AIDS 97
Figure 4.7 Natural Neighbour, Inverse Distance Weighting and Ordinary Kriging Maps 99
Figure 4.8 factors contributing to the spatial distribution of HIV/AIDS 111
Figure 4.9shows spatial distribution of HIV/AIDS from the previous study and the current study 112

Abbreviation

ACHAP	African Comprehensive HIV/AIDS Partnerships
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral
ARV	Antiretroviral treatment
BAIS	Botswana Aids Impact Survey
CDC	Centre for Disease control
CSO	Central Statistics Office
EAs	Enumeration Areas
FSW	Female Sex Workers
GIS	Geographical Information System
GPS	Geographical Positioning System
HCT	HIV Counselling and Testing
HIV	Human Immunodeficiency Virus
MPT	Medium Term Plan
MWM	Men Who Have Sex with Men
NACA	National Aids Coordinating Council
NCI	United States National Cancer Institute
OR	Odds Ratio
PEPFAR	President's Emergency plan for AIDS Relief
PMTCT	Prevention of Mother to child Transmission
SSA	Sub- Saharan Africa
STI	Sexual Transmitted Infection
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization

ABSTRACT

Background

The HIV/AIDS epidemic poses a serious challenge worldwide and threatens human welfare. The severity of this epidemic varies from district to district in Botswana, with the highest prevalence recorded in Selebi-Phikwe and North-East districts. This study seeks to provide an update of the spatial distribution of HIV/AIDS prevalence using different interpolation procedures. The study also seeks to identify socio-economic, geographic, demographic and behavioural risk factors that promote the spatial distribution of HIV/AIDS prevalence.

Data and Methods

This study used secondary data from the Botswana AIDS Impact Survey IV (BAIS IV), a nationally representative sample survey conducted between January and April 2013. The respondents of this study were 116482 HIV positive individuals aged 15-49 years in 12 selected districts. Inverse distance weighting, kriging and natural neighbour interpolation methods were used within ArcGIS Geographic information systems (GIS) software to generate continuous surfaces of HIV/AIDS prevalence. Spatial autocorrelation and clustering of HIV prevalence were analysed using Moran's I and Getis-Ord General G statistics. Local indicator of spatial association (LISA), Getis-Ord Gi* and Kulldorff scans were employed to identify districts that had high or low concentration of HIV/AIDS (Hot/Cold spots). Logistic regression was used to identify factors that were associated with spatial distribution of HIV/AIDS prevalence.

Results

Overall HIV/AIDS rates are high with Selebi-Phikwe having the highest of 18.6% followed by Francistown and Central-Mahalapye with 15.7% and 13.8% respectively. Females have a higher prevalence rate (62.7%) than men (37.3%). HIV/AIDS was also observed to be higher among the unmarried (47.7%), Christians (82.6%), fulltime workers (40.5) and among those with junior education (44.9%). Moran's I and Getis General G statistics revealed that HIV/AIDS is spatially distributed with values 0.135, p= 0.0481 and Z = 24101. P =0.016

respectively. Central-Serowe district was identified as the hotspot by both Kulrdoff scan and Getis Ord with a Log likelihood of 11248.11 and relative risk of 7.6. Three secondary clusters were also identified and these are Selebi-Phikwe, Francistown and Central-Mahalapye with relative 1.36, 1.16 and 0.28 respectively. On the contrary, the results revealed Ngwakwetse and Kgalagadi north and south as cold spots. Ordinary kriging with RMSE (6.3263) was found to be the best interpolation method and the continuous maps indicated that HIV/AIDS is concentrated in the south, northeast and the central districts. The logit model showed that alcohol, the number of sexual partners and condom use are the common risk factors contributing to the spatial distribution of HIV/AIDS in the selected districts of Botswana.

Conclusion

The spatial differences of HIV/AIDS across the selected districts and the identification of hot/cold spots suggest that a one size fits all kind of intervention might not be suitable for implementation in the different districts. Intervention should therefore, incorporate spatial variability and the identified risk factors. Reduced logistic regression model was significant in identifying factors associated with HIV/AIDS.

KEY WORDS: Spatial distribution, Risk factors, Interpolation, Demographic, HIV/AIDS Botswana Aids Impact Survey IV (BAIS IV), Distribution

CHAPTER 1

STUDY ORIENTATION

1.1. Introduction

Acquired immune deficiency syndrome (AIDS) is an infectious disease caused by the human immunodeficiency virus (HIV). There are two types of the HIV virus, HIV-1 and HIV-2, both of which ultimately cause AIDS. AIDS is one of the most devastating public health problems worldwide. The first case of AIDS was recognized among homosexual men in the United States in 1981 (Jeefoo, 2012). Since its first identification three decades ago AIDS has infected at least 60 million people and caused more than 25 million deaths (sharp and Hahn, 2011). Countries in the Sub-Saharan African (SSA) region are more heavily affected by the HIV/AIDS epidemic compared to other regions. The pandemic had far fetching effects and these include decreased life expectancy, increased mortality rates and the social and economic burden of orphan hood.

These countries have an estimated 22.4 million people living with the HIV (UNAIDS 2010), with more than two thirds of global cases of HIV/AIDS. This implies that for every three individuals affected by HIV globally, two live in the SSA region. About 1.4 million people have died due to Acquired Immune Deficiency syndrome (AIDS) related diseases. Furthermore, in 2010, new HIV infections in the region were estimated at 1.9 million (UNAIDS, 2010). The adult population's HIV prevalence rate in the region was estimated to be 5.2%, while the global prevalence was about 0.8% during 2010. There is considerable variation of HIV prevalence between sub-regions. Western and central parts have comparatively lower prevalence rates (2%) compared to the southern parts, whose prevalence rates range between 15% and 30% (UNAIDS, 2010).

The epidemic in Africa differs from that elsewhere in the world in that it is mostly spread through unsafe heterosexual intercourse, which means that women are more heavily affected in Africa than in other regions. This has two significant implications. Firstly, there is a massive vertical transmission to infants and secondly, if one parent is infected, the other is also likely to become infected, raising the number of orphans (Daniel, 2000: Barret, 2007).

According Barret, (2007) the result of the epidemics in sub-Saharan Africa has been the loss of a whole generation of young adults, who leave over 12 million orphans. 005). The peak ages of AIDS cases in sub-Saharan Africa are 20-29 years for females and 25-34 years for men (Barret, 2007).

1.2. Background of the study

The beginning and terrifying spread of the HIV and acquired immune deficiency syndrome (AIDS) presents a serious challenge and threatens the overall human welfare. Botswana is experiencing one of the most severe HIV/AIDS epidemics in the SSA region and worldwide. The national HIV prevalence rate among adults is the highest (23.4%), which is the second highest in the world, behind Swaziland (26%) (UNAIDS, 2012). The 2013 Botswana AIDS Impact Survey IV (2013 BAIS IV) data indicated a national HIV prevalence rate of 18.5 %, while the rate of new infections (incidence) estimated was 2.61%.

Figure 1.1 HIV/AIDS Prevalence Rate by Age and Gender (Adapted from BAIS IV)

The national HIV prevalence rate was found to be at its peak between the ages 35-39 years (43.7%) and 45-49years (41.8%). Figure 1.1 show the prevalence rate by age and gender The prevalence patterns among females show some disparities, with female prevalence rising to nearly 50.6% at an earlier age whilst that of male rises to 43.8% in the 40-44 years age group.

Female prevalence is higher than for men of the ages below 50 years, while male prevalence is higher for those beyond 50 years (Central Statistics Office, 2013).

Individuals, families and communities are suffering because of the impact caused by HIV/AIDS. There is a critical geographic variety in the regional level and HIV prevalence is varied with the most noteworthy recorded in Selebi-Phikwe, furthermore in the north East districts having a prevalence of 41.6% (Kandala *et al.* 2012). The high prevalence among teenagers and adults (ages 15-49) has serious implications and consequences for Botswana's economy and well-being in general.

Kandala *et al.* (2012) studied the geography of HIV/AIDS. The study produced age- and location-adjusted prevalence maps and these are important when focusing on of HIV instructive projects. For effective intervention and prevention of the epidemic, there is need for understanding both its spatial patterns and factors promoting its spatial distribution. In Botswana, no study combined spatial distribution and factors associated with HIV/AIDS. Hence the country keeps on being essentially tested in addressing the issues surrounding the prevention the transmission of the epidemic as signalled in the BAIS IV overview report (Botswana AIDS Impact survey, 2013 (BAIS IV). The current study uses 2013 BAIS IV data to investigate the spatial patterns of the pandemic in Botswana. The study further determines the social, cultural, economic, behavioural and other distal factors that might explain the spatial distribution of the epidemic. Understanding the spatial distribution of the epidemic and the factors promoting it, could help health officials to determine where intervention and prevention programmes are needed most.

1.3. Importance of Spatial Analysis

The spatial component of health data can play a crucial part in helping explain variability in risk because health status, environmental hazards, population numbers, demographic and socioeconomic profiles, and other relevant characteristics (e.g., susceptibility and exposures) all vary across space.(Pfeiffer et al,. 2008). Spatial analyses are important to health research because they highlight concepts of proximity and access, isolation or exposure, neighbourhoods' effects and boundaries, and diffusion (Logan, 2012).According to Chimoyi and Musenge, 2014 HIV/AIDS has a geographical structure that determines its epidemiology, a characteristic of spatially correlated data. Transmission in neighbouring locations is influenced by common spatially correlated and this results in the creation of spatial heterogeneity of diseases on a community, regional or national level (Chimoyi and Musenge, 2014). "Spatial analysis therefore takes into account these variations providing parameter estimates and predictions that can be used to produce spatial risk maps with the outcomes of interest in areas otherwise not sampled ",(Chimoyi and Musenge, 2014 p2). Hence, the use of spatial rather than standard regression models is suitable for accounting for these variations at district level in Botswana. The methods also allow examination of the spatial heterogeneity and identification of hotspots of diseases independently of administrative boundaries and spatial heterogeneity in the assessment of risk factors. The use of spatial analysis has provided important data information to national health policy makers for developing effective interventions and allocation of finance and human resources based on the local situation (Barankanira et al., 2016). Once cluster detection has been accomplished, then hypotheses and testable explanations can then be generated which attempt to give insight as to the cause of these patterns

1.4. Problem Statement

The government of Botswana has put in place various programs and policies to combat the spread of HIV. Some of them are the prevention of mother-to-child transmission (PMTCT), antiretroviral therapy (ARV), prevention of sexual transmission, HIV counselling and testing (HCT), sexually transmitted infections (STI) management, preventing Blood Borne Transmission (BBT), and many more. These interventions have been scaled up and the coverage rates have been above average. Currently in Botswana antiretroviral treatment has an estimated coverage of 87% of all patients in need of ART (UNAIDS, 2013).

The issue is that regardless all these efforts, Botswana keeps on being constantly tested in addressing issues encompassing the prevention of HIV/AIDS transmission. Equally noteworthy is that still after 12 years of a well-resourced HIV/AIDS response and in a country of only 2.2 million people, the prevalence is still high. In addition, substantive information about HIV/AIDS remains low in more than half of the young adults. The present status of the pandemic is such that Botswana is among nations in Southern Africa with the most elevated burden of HIV having HIV prevalence rate of 18.5% (UNAIDS, 2013). The BAIS IV 2013 summary results show that there is an increase of 0.9% compared to BAIS III 2008 in the prevalence rate.

Furthermore, there is little research on the spatial distribution of HIV/AIDS in Botswana using different interpolation techniques. Numerous studies internationally have mostly concentrated on aspects such as gender role attitude (Letamo, 2011), the impact of circumcision (Ayiga and Letamo 2011), the efficacy of therapies and vaccine development (Koff and Berkely, 2010), exploring the contribution of other diseases on HIV transmission (Van Houdt *et al.*, 2010) and the geography of HIV/AIDS (Kandala *et al.*, 2012). However, limited research on the spatial distribution has been conducted in Botswana using geographical information systems to better comprehend the spatial epidemiology of HIV/AIDS.

1.5. Rationale of the study

This study considers the importance of the spatial distribution of HIV/AIDS in Botswana. This is because studies have shown that HIV/AIDS is not equally distributed in the country. HIV/AIDS has shown to be higher in Selebi-Phikwe, Sowa and Francistown (Kandala *et al.*, 2012). This study seeks to identify the hotspots or cold spots of the epidemic in the country and to also identify risk factors that contribute to the spatial distribution of the epidemic. Understanding the variation of the epidemic and the risk factors associated with it can help to ensure that scarce resources are allocated efficiently to the districts where there is a greater risk.

The findings from this study will furnish significant information to the government and policy makers on the spatial distribution of the epidemic in the country. The Ministry of Health which is directly responsible for health delivery services in terms of policy formulation, implementation, monitoring, evaluation and regulation of health delivery services stands to benefit from the outcome of this study. The ministry and other organisations would be informed about districts that need more intervention and the risk factors that need urgent attention. The National AIDS Coordinating Agency and Non-Governmental Organisations will benefit from this research in terms of having knowledge about districts that need more HIV/AIDS intervention resources.

1.6. Aim of study

The aim of the study is to use current data to assess the spatial distribution of HIV/AIDS in Botswana using inverse distance weighting, natural neighbour and ordinary kriging. Additionally it determines HIV/AIDS prevalence risk factors associated with the spatial distribution. The study also aims to produce maps to identify high and low risk areas and compare them with previous spatial distribution maps. The study further aims to identify hot/cold spots and factors promoting the spatial patterns using 2013 BIAS IV data.

1.7. Research Questions

The main research questions for this study are stated as follows:

1.6.1 What are the similarities or dissimilarities of HIV/AIDS characteristics in districts sharing a common border?

1.6.2 Which districts have high or low concentration of HIV/AIDS?

1.6.3 Which is the best spatial interpolation method for BAIS IV data?

1.6.4 What factors promote the spatial patterns of HIV/AIDS in Botswana?

1.8. Research Objective(s)

The study is set to achieve the following objectives:

- To determine if districts have similar or dissimilar HIV/AIDS characteristics
- To classify districts according to a high or low HIV/AIDS prevalence rate.
- To construct continuous surface maps of HIV/AIDS prevalence.
- To determine the best spatial interpolation method for analysing spatial distribution using BAIS IV data.
- To provide suggestions on improving HIV/AIDS interventions.
- To determine factors that promotes the spatial patterns of HIV/AIDS.

1.9. Significance of the Study

Botswana is rated second in the world among countries with the highest HIV infection rate, with one in three adults infected by the epidemic (UNAIDS, 2012). Several studies worldwide have been undertaken on the spatial distribution of HIV/AIDS using different spatial interpolation methods with demographic and health survey data (DHS). Very little research has been done that compares different spatial procedures in the spatial analysis of HIV/AIDS data. The only research on the spatial distribution of HIV/AIDS epidemic in Botswana was conducted by Kandala *et al*. (2012) using 2008 Botswana Aids Impact Survey III. The Bayesian geo-additive mixed model based on Markov Chain Monte Carlo was utilised by the researchers to map the geographic distribution of HIV prevalence in the 26 districts.

Also, several studies have been carried on different risk factors of HIV/AIDS Greener et al., (2000) looked at influence of HIV/AIDS on poverty and inequality, Letamo (2011) at gender role attitude on the epidemic and Ayiga and Letamo (2011) on the impact of circumcision on the epidemic Although the studies are helpful, they lacked the spatial component of the disease. Appreciating the spatial variation of the epidemic infection in a country together with its drivers is crucial for establishing where prevention and treatment programmes need to be focussed because of the scarce resources. There is little research on spatial distribution of HIV/AIDS in Botswana which used different interpolation techniques.

A study that combines BIAS data with socio-economic, geographical and cultural factors to investigate factors that contribute to high risk of infection and how and why certain districts of Botswana have high prevalence rates than others is needed. The current study seeks to explore the spatial distribution of HIV/AIDS using different interpolation methods (inverse distance weighting (IDW), kriging and nearest neighbour). Likewise, studies that also compare the performance of these interpolation methods using BIAS IV data and determining factors that influence the epidemic's spatial patterns are needed. Furthermore, the current study analyses the influence of spatial autocorrelation patterns of nearby districts, created spatial maps and compared these with previous maps.

This study might be beneficial to the body of the literature on spatial distribution by using different methods for analyzing spatial distribution using demographic and health survey data. It will also contribute to literature by providing current statistics on HIV/AIDS and will help government, policy makers and service providers to identify areas that need intervention. Furthermore, this study will benefit Botswana government by showing the recent geographical patterns of HIV/AIDS in Botswana and providing information as to why some districts of Botswana have high prevalence rate than others. Knowledge of this information might help government, policy makers and service providers in planning, monitoring and evaluation of health programmers. This may also help the government of Botswana to restructure its intervention programs for each district and come up with specific programs that will prevent the spread of the epidemic

1.10. Assumptions

This study is based on the belief which states that everything is related to everything else, close things are more related than distant things (Tobler, 1970). Statistics aims to characterise a population based on a sampling of that population. If samples are not randomly selected, the sample population might be biased and calculations from a biased sample population will not accurately describe the population of interest, hence the assumption of the study is that samples are randomly chosen from the population. The study also assumes that HIV/AIDS prevalence is homogeneous in Botswana districts.

1.11. Outline of the thesis

The study is composed of five chapters. The first chapter gives the introduction of the epidemiology worldwide in sub-Saharan Africa. Chapter two gives the literature review of HIV/AIDS worldwide, in Africa, in sub-Saharan and in Botswana. Firstly it looks at prevalence and modes of transmission. Secondly it examines the uses and importance of GIS in health research. Thirdly it examines spatial analysis, its importance and where it has been utilised. Lastly, it looks at studies done using different interpolation techniques. Chapter three describes the research methods employed in this study. Emphasis is on autocorrelation methods, hotspot detection methods, interpolation techniques and logistic regression. Chapter four reports on the findings of the study and discusses their implications. Chapter five looks at discussions, contribution of the study and recommendations.

1.12. Definition of terms

This section gives a definition of the terms frequently used in the study.

Spatial Analysis: Analytical technique which accounts for spatial variations inherent in spatial data which can be used for statistical inference (Graham et al., 2004)

Clustering: Grouping of health events situated closely together in relation to time and/or space (Graham et al., 2004).

GPS (Global Positioning System): A device that collects spatially distributed data in real time (Moodley, 2010).

Demographic and Health Surveys: These are nationally- representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health and nutrition (Otwombe, 2013)

GIS (Geographical Information System): This is a series of tools for the acquisition, storage, retrieval, analysis and display of spatially referenced data (Graham *et al.*, 2004; Moodley, 2010).

Prevalence: The actual number of cases alive with the disease either during a period (period prevalence) or at a particular date in time (point prevalence. (Shields and Twycross, 2003).

Incidence: This is number of new (or newly diagnosed) cases of a disease occurring during a period of time (Shields and Twycross, 2003).

1.13. Conclusion

This chapter outlined the background of HIV/AIDS in Botswana. It briefly provided prevalence rate of the disease in Southern Africa. The chapter also presented the aim of the study, objectives, research questions, problem statement, and significance of the study and the assumptions of the study.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

In Chapter 1, the study was oriented by looking at the general prevalence of HIV/AIDS, the aims, objectives and significance of this study. This chapter examines the literature review on the prevalence of HIV/AIDS. First the study looks at the statistics of HIV/AIDS worldwide, in Africa and in Botswana. It also examines literature on the importance of GIS and spatial statistical analysis in health research. The chapter ends by looking at literature on interpolation methods.

2.2. HIV/AIDS Worldwide

Research on HIV/AIDS has steadily increased since its discovery two decades ago. A wide range of topics have been researched since the discovery of this pandemic. These range from demographic implication of the study, research into intervention, best practices that may stop the spread of the disease and spatial distribution of the disease.

HIV/AIDS is a worldwide problem that has affected many people in the whole world. Since the beginning of the pandemic in 1981, 78 million people have been infected and about 39 million have died of the epidemic. There were 35 million people living with the epidemic globally in 2013 and of these, 3.5 million were children under the age of 15. In the same year 2.1 million people were newly infected (UNAIDS 2013). Of the 39 million people who have died due to the epidemic, 1.5 million have died of AIDS related diseases. These include pneumonia, herpes simplex, cancer, diabetes and tuberculosis which is the most common opportunistic infection related to the epidemic and the principal reason for death among individuals with the epidemic. These are called "opportunistic' diseases because they take advantage of the ill peoples' weakened immune system and they can cause destructive illnesses (Avert, 2014).

The burden of HIV/AIDS varies significantly between countries and regions. Sub- Saharan Africa (SSA) is the mostly affected region and it accounts for almost 70% of the people living with the epidemic worldwide despite that only 13% of the world's population live in SSA (Morison, 2001, Kandala *et al.*, 2012). About 88% of children living with the epidemic reside in this region. The national prevalence rate of most countries in sub-Saharan Africa is greater than 1% with South Africa having the most number of individuals living with the epidemic around the world (6.8 million). Swaziland (27.4%) has the highest prevalence rate in the globe (UNAIDS, 2015).

Nearly 2 million people in Latin America and the Caribbean are estimated to be living with HIV/AIDS and newly infected individuals were about 100 000 in 2014 (UNAIDS, 2015). The Caribbean region is the second hardest hit region after South Africa with an adult prevalence rate of 1.1% In this region, Haiti has the highest the epidemic prevalence rate (1.9%) and Brazil has the highest number of persons living with the pandemic ranging at about 610 000 -1 000 000 (UNAIDS, 2015). About 1.5 million people in Eastern Europe and Central Asia are estimated to be living with the epidemic and this includes 140 000 who are newly infected. In this region, the Russian Federation and Ukraine account for 85% of people living with the epidemic (UNIADS, 2015).

Worldwide, Brazil is one of the few countries that effectively checked the spread of the epidemic (World Health Report, 2004). Its primary HIV/AIDS programme was introduced in the State of Sao Paulo in 1983. This occurred immediately upon realisation that four cases of HIV/AIDS were been reported. The structure and role of the public health system in Brazil has influenced significantly its response to the epidemic. By the end of 2002, about 260 000 cases had been reported to the Ministry of Health. The World Bank had estimated that the prevalence rate of the epidemic in Brazil by 2000 would be 1.2%, but due to the country's effective response programme, the prevalence was half of the World Bank's prediction (0.6%). A large-scale widespread antiretroviral distribution programme was first implemented in Brazil. And about 130 000 people in Brazil are now provided with free drugs for opportunistic infection (World Health Report, 2004).

Figures 2.1, 2.2 and 2.3 show adult prevalence rates, the number of people living with HIV/AIDS and the number of new infections worldwide for the year 2014.

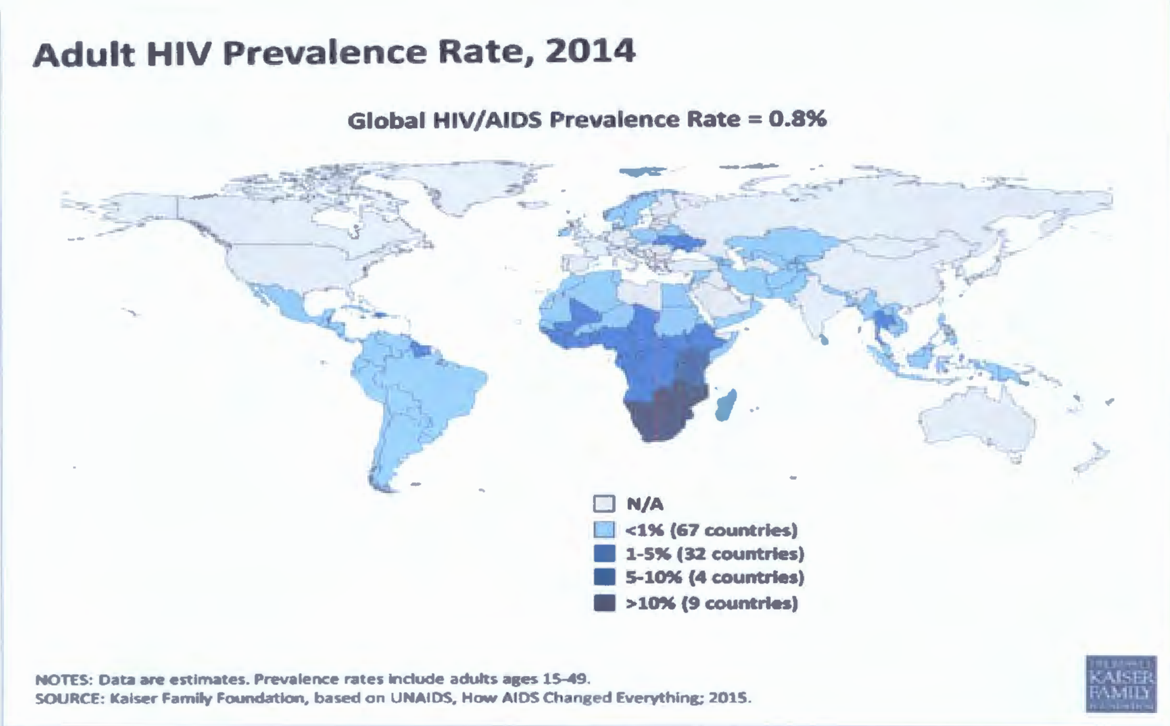


Figure 2.1 Adult HIV/AIDS Prevalence Rate for the Year 2014 (Source: Kaiser Family Foundation, 2015).

Figure 2.1 shows the prevalence rate of HIV/AIDS for the years 2014. The darker blue colour represents a high prevalence, the lighter blue colour lower prevalence and grey colour for countries no records. Southern Africa is severely affected HIV/AIDS with nine countries having a prevalence rate greater than 10%.

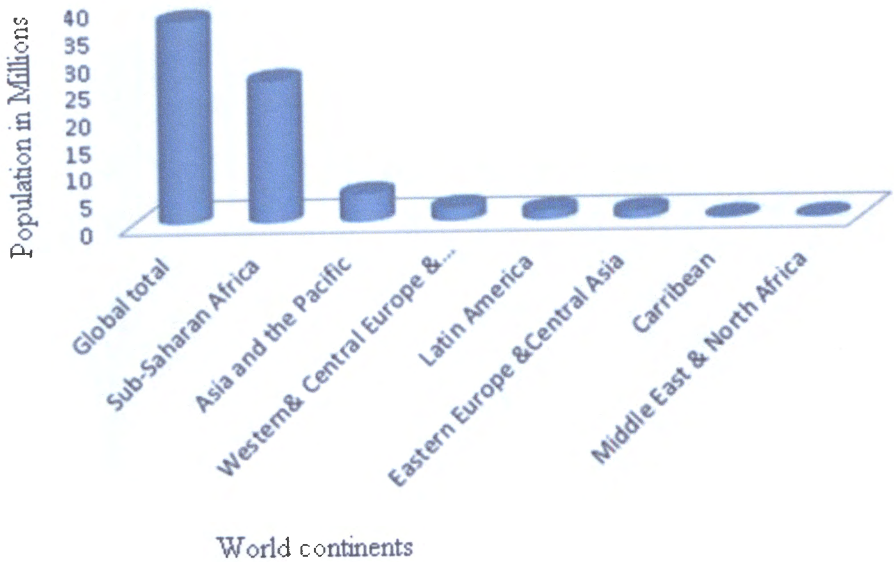


Figure 2.2 Total number of people living with HIV/AIDS for the year 2014 (Source: Kaiser Family Foundation, 2015)

Globally Sub-Saharan Africa has the highest number of people living with HIV/AIDS followed by Asia and the Pacific and Eastern Europe and Central Asia as seen in Figure 2.2. The Caribbean, Middle East and North Africa have the least number of people living with HIV/AIDS.

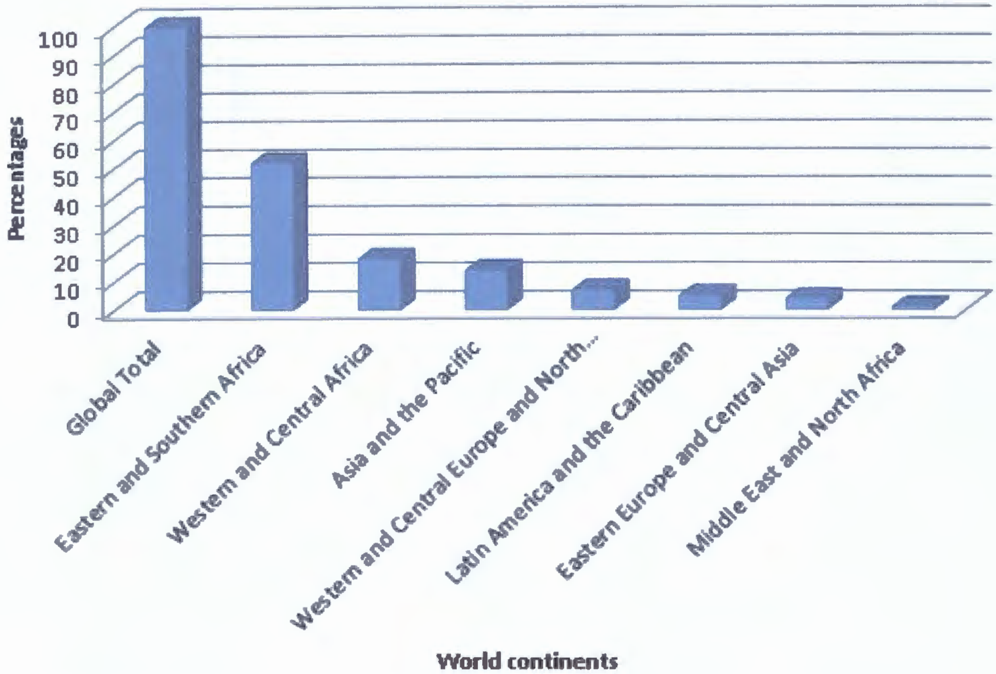


Figure 2.3 Number of new infection for the year 2014 (Source: Kaiser Family Foundation, 2015)

In Figure 2.3 Sub-Saharan Africa has the highest number of new infections globally followed by Asia, the Pacific, Western and Central Europe. Numbers of new infections are least in the Caribbean, Middle East and North Africa.

2.2.1. Modes of transmission

HIV/AIDS can be transmitted through different modes and these include sexual transmission, parental, use of contaminated objects and mother-to-child transmission (Morison, 2001).

2.2.2. Sexual transmission

Most of the people infected by HIV/AIDS since the pandemic began have caught the virus through either sexual transmission, parental or mother-to-child transmission. The most common mode of the epidemic transmission globally is sexual transmission (Hladik and McElrath, 2008; Cohen and Galvin, 2004). In females, first sexual intercourse maybe associated with high transmission probabilities of pandemic. In sexual transmission, receptive anal intercourse has been found to be more risky than vaginal intercourse and it is the major form of transmission among men who have sex with men (Morison, 2001).

2.2.3. Parental transmission

Parenteral transmission is a kind occurring outside of the alimentary tract, such as in subcutaneous, intravenous, intramuscular, and intrasternal injections according to (Berkely, 1991). The author posits that in Africa, heterosexual transmission has been discovered as the primary mode of HIV infection. However the use of large quantities of injections by health care personnel and traditional healers both in and out of the health care setting indicates that parenteral transmission could also be an influencing factor to HIV infection in the region (Berkely, 1991). Discoveries further reported that parenteral transmission can also occur by the transfusion of infected blood.

2.2.4. Mother-to-child transmission

It is believed the about 90% of HIV/AIDS transmissions are through mother-to-child transmission. About two-thirds of the transmissions happen in utero and at delivery, while a third of the transmissions occur through breast-feeding. Mother-to-child transmission is estimated to be about 5.1 million (Morison, 2001). In Africa, countries like Botswana, Mozambique, Namibia, South Africa, Swaziland and Uganda have succeeded in meeting the worldwide plan target of bringing down mother-to-child transmission by 90% (UNAIDS 2016). However, there are some countries that are still facing major challenges of rolling out effective PMTCT services. New HIV infection among children has reduced by about 40% in Angola, Cote d'Ivoire and Nigeria since 2009 (UNIADS, 2016). Figure 2.4 shows the prevention-of-mother-child progress against Global plan targets in selected priority countries in Africa.

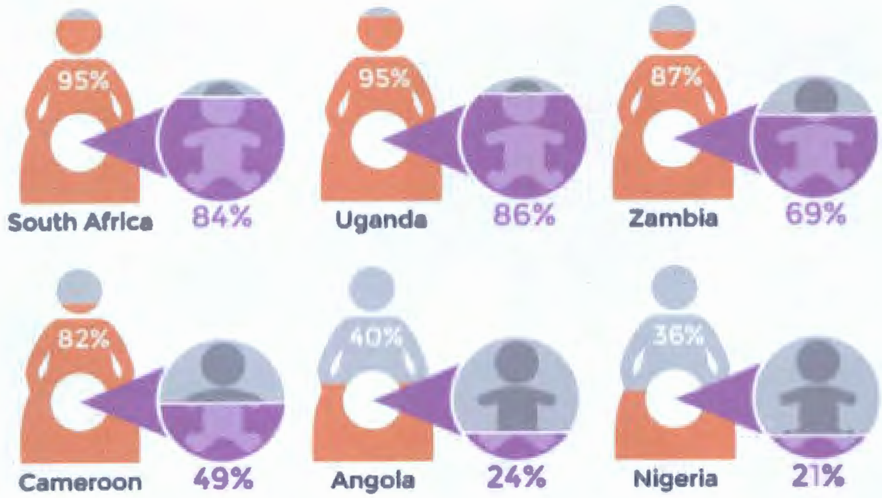


Figure 2.4 Prevention progress of PMTCT in selected priority countries in Africa. Source: UNAIDS 2016. On the fast track to an AIDS free generation

2.3. HIV/AIDS in Africa

In Africa, HIV/AIDS is one of the most significant public health concerns of our time, and perhaps, in the history of mankind and it is one of the top causes of death (UNAIDS 2016). The population of people living in Africa is slightly less than 15% of the total population of the world and yet Africans account for nearly 70% of those who live with HIV and are dying of AIDS (Essex *et al.*, 2007). The first HIV/AIDS case in Africa is believed to have occurred in Kinshasa, the capital city of Congo, in 1970. The virus which was brought by a traveller from Cameroon to Congo by the river entered a wide urban sexual network and quickly spread. This marked the first heterosexually spread of HIV/AIDS epidemic in Africa (Avert, 2014).

HIV spread rapidly in east Africa in the 1980s and it became more destructive than in West Africa. The pandemic was fast-tracked by wide spread labour migration, high rate of men in the urban population, low status of women, absence of circumcision and sexually transmitted diseases. In East Africa, especially in Nairobi, 85% of the sex workers were infected by HIV/AIDS in 1986. Sex workers contributed immensely to the spread of the epidemic in this region. According to Avert (2014), the epidemic spread to western Equatorial Africa and western African nations in the early 1980s. Uganda was hard hit by the pandemic in the 1980s. The virus did not cause much harm in western Equatorial countries of Gabon, Congo-

Brazzaville and Cameroon. The western Equatorial countries were not hard hit by the pandemic because of long distances between the cities also there is a lot of violence and insecurity. The rapid spread of the epidemic was facilitated by truck drivers alongside other migrants such as miners, traders and soldiers who engaged with sex workers during their travel. As indicated earlier, Uganda was hit hard by the pandemic, hence it had 35% of its truck drivers and 30 % of its military personnel testing positive for HIV (Avert, 2014).

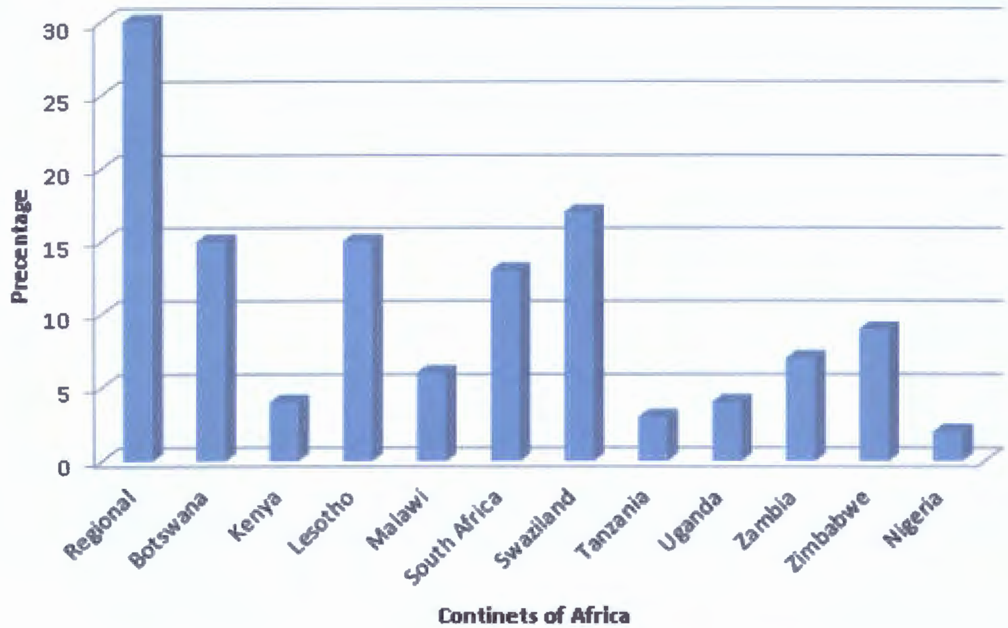


Figure 2.5 Number of people living with HIV/AIDS for the year 2014 (Source: Kaiser Family Foundation, 2015)

Figure 2.5 shows that South Africa has the highest number of people living with HIV/AIDS followed by Nigeria. Botswana, Lesotho and Swaziland have the least number of people living with the pandemic.

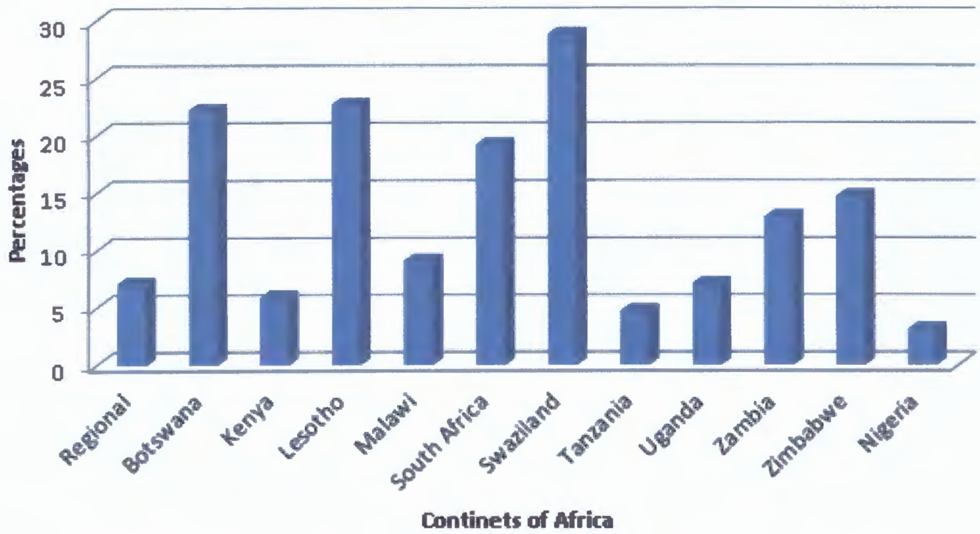


Figure 2.6 Adult Prevalence Rate for the year 2014 (Source: Kaiser Family Foundation, 2015)

Almost all the countries' shown in Figure 2.6 have high adult prevalence rate. Nigeria, Tanzania and Kenya have low adult prevalence rates. Swaziland has the highest adult prevalence rate followed by Botswana, Lesotho and South Africa.

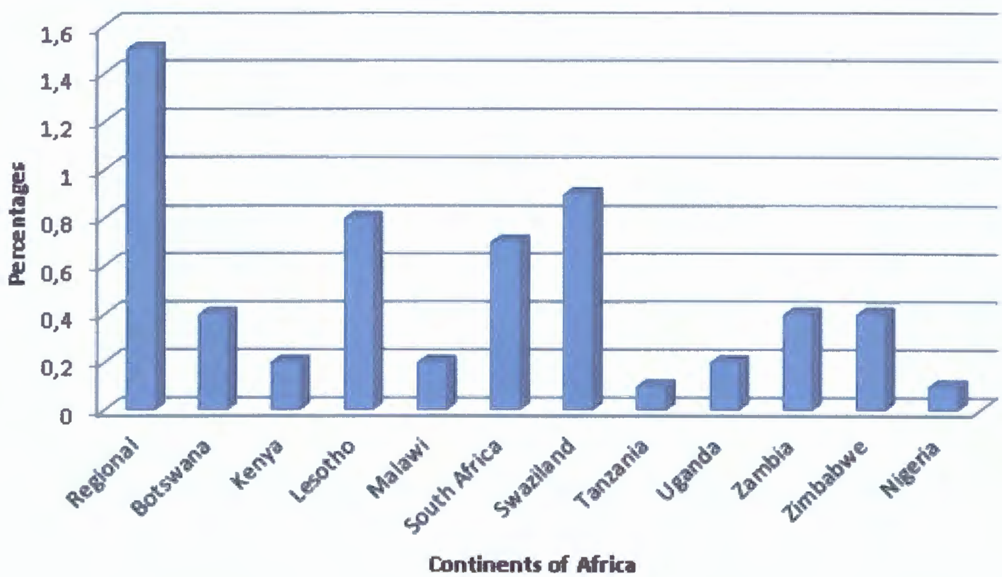


Figure 2.7 Number of New Infections for the year 2014 (Source: Kaiser Family Foundation, 2015)

In Figure 2.7 South Africa has the highest number of new HIV/AIDS infections regional followed by Nigeria, Botswana and Uganda. Lesotho and Swaziland have the least number of new HIV/AIDS infections regionally.

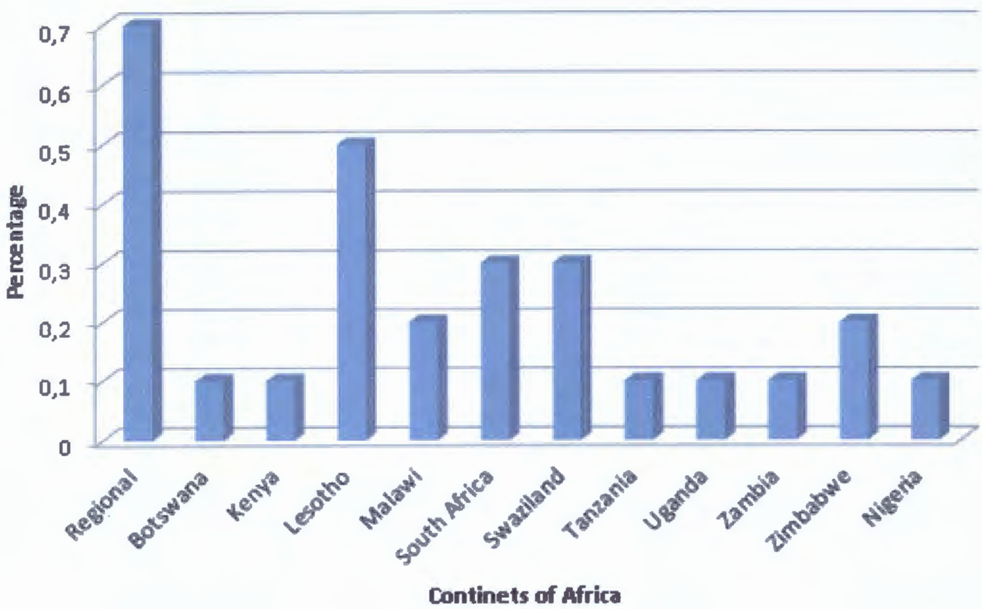


Figure 2.8 Number of Deaths for the year 2014 (Source: Kaiser Family Foundation, 2015)

Figure 2.8 shows that the number of HIV/AIDS related deaths for the year 2014 are higher in Nigeria and South Africa and lower in Botswana, Lesotho and Swaziland.

In Southern African countries the virus arrived moderately late, but it had devastating effects on the general population. Before the end of the 1980s, the HIV pandemic in the southern African nations of Malawi, Zambia, Zimbabwe and Botswana were nearly surpassing East Africa (Avert, 2014). Malawi had 980 000, Botswana 350 000, Zambia 1.2million and Zimbabwe 1.4 million people living with HIV/AIDS. To this day Southern African countries remain the most affected by the HIV/AIDS epidemic. The region constitutes only about 2% of the world's population, yet worldwide 34% of individuals living with HIV/AIDS live in these countries (UNAIDS 2010).

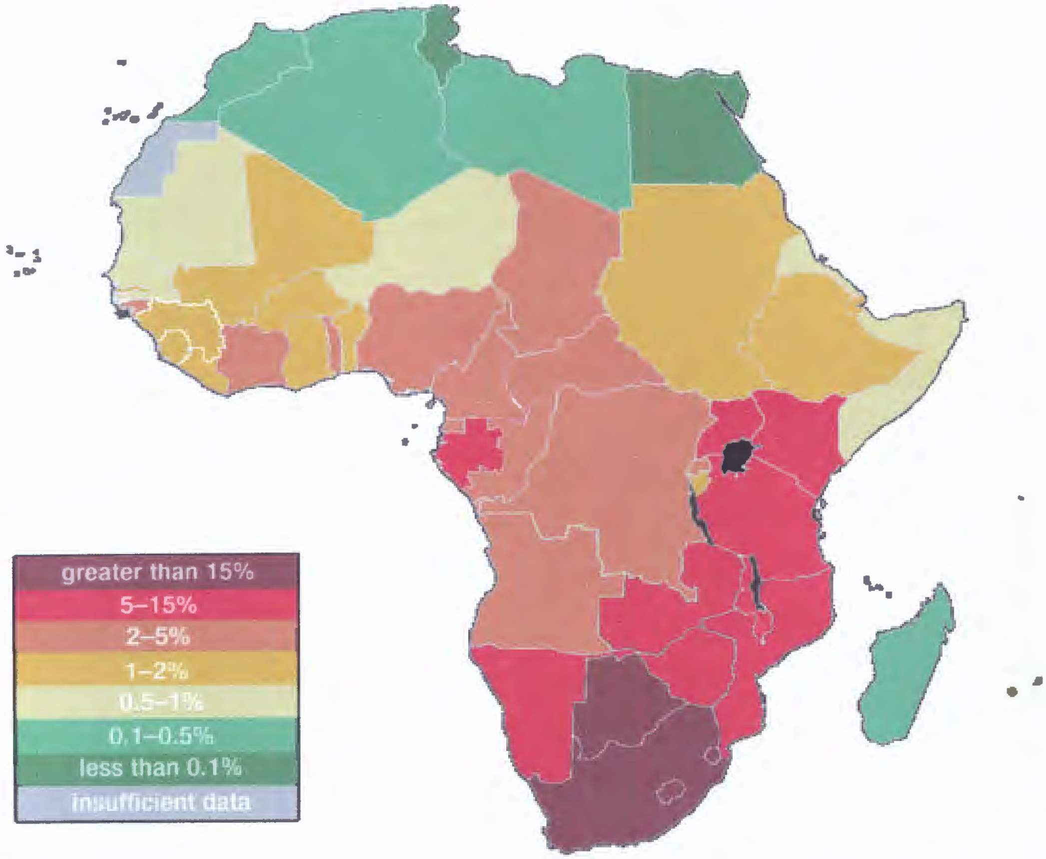


Figure 2.9 Statistics and facts about HIV in Africa (Adapted from: AIDS in Africa, 2013. Medwiser

2 3.1. Impact of HIV/AIDS in Africa

The HIV/AIDS epidemic has had a number of impacts in the Sub-Sahara region, most obvious impacts being sicknesses and the number of lives lost. There were about 1.6 million new HIV infections and 1.2 million AIDS-related deaths in 2012 (UNAIDS, 2013). The epidemic have impacted significantly upon the education sector, labour and productivity and the wider economy. Sub-Sahara Africa has scaled up its antiretroviral treatment (ART) across the region and this has resulted in decreased annual number of new infections by 34% since 2001. The number of people receiving ART has increased from 56% in 2011 (UNAIDS 2012) to 68% in 2012 (UNAIDS 2013).

Poverty has been linked to the spread of HIV/AIDS in Sub-Sahara Africa, but the association is very complicated and research remains unconvincing. There are 48.5% of people in sub-Saharan Africa who are surviving lower that the poverty line ($1.25 a day) (UNAIDS, 2012). Poverty can drive people to desert their homes in order to search for work making them vulnerable to exploitation especially, women who may be forced into early marriages and some into sex work (UNAIDS 2012). Poverty can affect people in many different ways, for example, the epidemic infected person can use all his/ her resources in treatment despite being economically unproductive due to their weak bodily condition.

The HIV epidemic has caused extreme and extended effects upon households in sub-Saharan Africa. Many families have lost their principal income earners, who either have died, or are too sick to work. This puts an overwhelming budgetary load on families who have to pay for ever increasing medical costs, forcing a significant number of them into poverty. As a result, many families have to provide home-based care, further lessening their earning capacity and placing more demands on their resources (Avert, 2014).

Hajizadeh *et al.* (2014) examined the socioeconomic inequalities in HIV/AIDS prevalence in 24 countries in Sub-Saharan Africa region. Their study exhibited that the epidemic was intense among wealth individuals in most of SSA countries. The only nations where the epidemic was intense among individuals living in poorer family units are Swaziland and Senegal. Among the poor in the urban areas and among wealthier adults in rural areas the epidemic was intense in Zambia, Kenya, Uganda and Lesotho. However, the stratified investigation demonstrated that the epidemic was generally concentrated among wealthier men and women.

Several studies conducted in sub- Saharan Africa found that there is a common relationship between young men and women which are associated with unsafe sexual behaviour and low condom use (Madlala, 2008). This increases their risk of contracting HIV/AIDS. Sex workers are a group that is also at high risk of the epidemic infection in sub-Saharan Africa with an average prevalence rate of 20% as compared to 3.9% globally (UNIADS, 2013).

Sub- Saharan Africa's life expectancy remained stagnate at 49.5 years between 1990 and 2000 (Joint United Nations Programme, 2013 (UNDP). In 2006, UNAIDS reported that 20 years of life expectancy had been wiped off by HIV/AID in many countries. This might have been generally attributed to child mortality, which is associated with an increase in the mother-to-child transmission during pregnancy (UNAIDS, 2006). The scaling up of antiretroviral treatment saw an increase in life expectancy by 5.5 years in the period 2000-2012. However, most countries have low life expectancy. In Swaziland and Lesotho life expectancy is equally low (48.9 years and 48.7 years) respectively. (Joint United Nations Programme, 2013 (UNDP).

Women and girls are becoming more vulnerable to the HIV/AIDS pandemic (Agyei- Mensah 2005). Young women between the ages of 15-25 have been found to be at a higher risk of the epidemic infection than their male counterparts (Agyei- Mensah, 2005). Gender relations in most countries are characterised by unequal balance of power between men and women, with women being deprived of opportunities of going to school, training, and income generating activities, property and health care services. Because of these factors, women are not in position to protect themselves from pandemic as well as being unable to access knowledge about health, treatment and care (WHO, 2004).

2.4. Factors affecting the spread

Several factor shave been reported to contribute to the spread of HIV/AIDS in Sub-Saharan Africa (Hoshi, 2016). Chimoyi and Musenge (2016) carried a research to identify risk factors associated with HIV/AIDS among young people age 15-24 years in Uganda. The study employed Maximum likelihood-based logistic regression models to explore the non-spatially adjusted factors associated with HIV infection. The findings identified marital status, sexual debut, sexual transmitted infection (STI) alcohol use and condom use as predictors of the epidemic. Seloiwe (2005) conducted a study among University of Botswana students to investigate factors influencing the spread of HIV/AIDS among University students The study findings revealed that alcohol, drug abuse, unprotected sex, frequent change in sexual partners, sex for financial gain, for prestige, for good grades and to relieve stress were factors contributing to the spread of the disease among the students.

In other studies Nyindo (2005) and Kalipeni et al., 2007 identified six major drivers of HIV/AIDS and these are historical context of colonialism, labour migration, gender, poverty, disease burden and government attitudes. Imelda and Kalipeni (2010) used sentimental data to examine spatial distribution of HIV/AIDS and factors that influence the pandemic in Zambia. Employing Ordinary Least Squares, the study revealed that literacy rates, unemployment, poverty and urban residence were risk factor of the epidemic in Zambia.

Various social and cultural traditions are some of the key issues that have reinforced vulnerability to HIV in Africa (Nandoya, 2014). According to Nandoya (2014) "religion prescribes ethical guidelines for many aspects of daily life and also navigates belief systems and norms surrounding sexuality." Many religions condemn the use of condoms and support a submissive role for women, foster gender inequality in marital relations, and promote women's ignorance in sexual matters as a symbol of purity (Nandoya, 2014). Marital status, early marriages and multiple sex partners are some of the socio-cultural factors that influence the spread of HIV/AIDS. In a study on sociocultural factors influencing the spread of HIV/AIDS in Africa, Nandoya indicates that gender inequality particular on sexual matters increases vulnerability to HIV transmission. Women and girls are more vulnerable to the pandemic; they do not have power to negotiate safer sex. Early marriage significantly increases the chance of HIV among young girls as they engage in sexual intercourse with much older, experienced and HIV positive husbands. Having multiple sexual partners is another socio-cultural factor that increases vulnerability to HIV infection. Cultural norms allow multiple sexual partners for men inside and outside marriage and this expose men and their partner to HIV infection (Nandoya, 2014).

2.5. HIV/AIDS in Botswana

Botswana has been severely hit by HIV/AIDS and is experiencing one of the most serious HIV/AIDS epidemics in the world. In 1985 Botswana's first case of the epidemic was reported in Selebi-Phikwe. From that time the epidemic has multiplied rapidly with pandemic prevalence levels reaching 36.2% in 2001, 33.4% in 2005 and eventually reducing to 31.8% in 2009 among pregnant women aged 15-49. The 2004 BAIS II and 2008 BAIS III indicated a national HIV/AIDS prevalence rate of 17.1% and 17.6% respectively. For the same period the prevalence rate for females and males was 20.4% and 14.2% respectively. The prevalence rate for 2013 BIAS IV was 18.5% which showed an increase of 0.9% compared to 2008 BAIS III.

(Central Statistics Office, 2013). High prevalence rates have been noted in the northern-eastern corridor of the country and the southern parts of the country had least prevalence rates.

The estimated prevalence for men, women and children under the age of 18 months is 19.2%, 14.2% and 2.2% respectively. There is a slight variation in HIV/AIDS between urban and rural centres. In urban centres pandemic is estimated at 17.5% while rural centres it's at 15.8 % (UNAIDS, 2015). HIV/AIDS has had a negative effect on the developmental gains that have been achieved by the country since its independence in 1966. These include economic growth, life expectancy and health care systems. In Botswana the primary mode of the epidemic transmission is heterosexual intercourse. The majority of young people in Botswana engage in sexual activities before marriage. Young women and the military are at a higher risk of HIV/AIDS infection as compared to other sectors of the population (Molatole and Thaga, 2006).

In spite of Botswana being seriously hit by HIV/AIDS, it was the first nation in sub-Saharan Africa to give free antiretroviral treatment to people living with HIV/AIDS. A few activities have been instituted by the Botswana government to attempt and battle the illness. In 1986 the purported Minimum Program was set up by the government under the study of disease transmission unit of the Ministry of Health. In 1987 a transient arrangement was produced to concentrate on developing national open attention to HIV. Around the same time the National AIDS Control Program was launched and it was aimed at creating short term medical responses.

The short-term plan was then trailed by the Medium Term Plan (MTP) for the Prevention and Control of HIV/AIDS from 1989 to 1993. The MTP gave approach and vital rules to activity since the origin of the National Aids Control Program. The MTP sketched out the role of the health sector and the Ministry of Health, with the support and help of different sectors and non-governmental organisations (NGOs) for HIV/AIDS counteractive action care and support. The second Medium Term Plan (MTPII) was along these lines received in 1997 to venture up endeavours in the battle against the pandemic. Its two fundamental objectives were to lessen HIV disease and transmission, and also to diminish the effect of HIV and AIDS at all levels of society in the nation. In 1998 Botswana was he first nation in Africa to give therapeutic prevention of mother-to-child transmission of HIV/AIDS. In 2002 the administration made a stride further and set up a HIV/AIDS National Strategic Framework (Avert, 2014).

Several HIV/AIDS prevention programmes have been rolled out since 1988. Some of these are the award winning teacher- capacity building programme which was aimed at improving the teacher's knowledge of HIV/AIDS and to reduce stigma surrounding the disease. Primary and secondary school were furnished with a television, video recorder, different resources and interactive HIV/AIDS education programme.

Mass media was utilised for the prevention of HIV/AIDS, particularly radio and television. The radio runs a drama called Mkgabaneng which has subjects that are connected to HIV/AIDS pandemic inside Botswana. It addresses issues such as HIV/AIDS treatment, loyalty, cultural traditions and facilities that are available (Avert, 2014). One of most effectively implemented HIV/AIDS programmes within Botswana's HIV/AIDS response is the prevention of mother-to-child transmission (PMTCT). Of the 11 000 pregnant women living with HIV/AIDS, 10648 (>95%) are on antiretroviral treatment. The PMTCT programme has reduced the mother-to-child transmissions rate to 2.49 % (Avert, 2014).

Another very successful programme launched for HIV/AIDS prevention intervention in 2002 in Botswana was the Masa treatment programme. Masa is a Setswana word meaning "new dawn," heralds the rising of a dawn over Botswana's struggle against the HIV/AIDS epidemic and promises Batswana the opportunity to live longer and healthier lives by giving people living with HIV/AIDS more time to nurture their families and to help build a better future for Botswana(Farahani *et al.*, 2014). The programme is widespread and it freely makes antiretroviral treatment accessible to all eligible citizens Farahani *et al.*, 2014). In 2013 an estimated total of 213 953 adults living with HIV/AIDS were receiving antiretroviral therapy. The coverage of children living with the diseases also increased to 84 % (Avert 2014: Farahani *et al.,* 2014).

2.5.1. Challenges

After making impressive HIV/AIDS response, Botswana is facing financial challenges to maintain its response. The success of Botswana's intervention programme was because of donor funding. However, many donors have either withdrawn or reduced their funding because Botswana is an upper- middle country. President's Emergency plan for AIDS Relief (PEPFAR) has reduced its funding by 30 million US dollars between 2009 and 2012, while Centre for Disease control (CDC) and African Comprehensive HIV/AIDS Partnerships (ACHAP) have withdrawn their financial funding for safe circumcision. The Gates Foundation withdrew its funding as well in 2013. The withdrawal of funding had negative implications on Botswana's national prevention and treatment programmes. The financial withdrawal has resulted in shortage of human resources (Avert, 2014).

2.5.2. Key affected populations in Botswana

There are a number of population sectors that are severely hit by HIV/AIDS pandemic in Botswana. Among these sectors are female sex workers (FSW) and men who have sex with men (MSM) who were for the first time included in the HIV/AIDS epidemic surveys in 2012 (Ministry of Health report, 2014). FSW in Botswana have a prevalence of 61.9 %(Avert 2014). FSW's vulnerability is increased as they regularly work in high risk situations where conditions change depending on the customer (UNAIDS, 2014). According to Merrigan, *et al.*, 2015, FSW can be paid more money for not using protection or at times are forced not to use protection by their clients, thus increasing their vulnerability to epidemic infections.

2.5.3. Determinants of HIV/AIDS in Botswana

There are several factors that tend to drive the spread of HIV/AIDS infection in Botswana. In this section socio- economic determinant, denial and stigma, socio-cultural determinants and mobility are discussed.

2.5.3.1 Socio-cultural determinants

Culture plays an important part in deciding the level of health of an individual, the family and the community. In Africa, the values of extended family and community significantly contribute to the behaviour of an individual (Airhihenbuwa and Webster, 2004). The position of women in the society, their limited power to negotiate issues of sex and less economic empowerment makes them vulnerable to HIV/AIDS infection. Most girls who engage in relationships with older men are ignorant and submissive, hence they are unable to discuss for safe sex (Letamo, 2003). Generally, women whether young or old, married or unmarried have no say over sex; men determine how they want sex. If a woman brings condoms at home, she is regarded as a prostitute.

2.5.3.2 Socio- economic determinants

HIV/AIDS generally affects the working age-group in most developing countries and is not just a medical issue but it is additional a socio-economic setback (Matshe and Pimhidzai, 2008). Economic hardship has resulted in poverty and lack of economic opportunities which has led young and single women to engage in unprotected sex in exchange for money and other basic services. This has led to increase the spread of the epidemic in the society. These economic hardships in rural areas have pushed young women to urban areas in search of employment, where they indulge in sex work for them to survive. Wealth and consumption patterns have been identified as contributors of the epidemic. People with high income exploit those with low income and exert unfair advantage in exchange of sex (Molatole and Thaga, 2006).

2.5.3.3 Regional trade transit point

As indicated earlier, Botswana is a landlocked nation with reasonably well-established transport structure. It is a transport centre for South Africa, Namibia, Zimbabwe and Zambia. These countries have high HIV/AIDS pandemic in Southern Africa (Molatole and Thaga, 2006). Botswana being a transit hub, there is a high number of people travelling from these neighbouring countries through it. The transit of people through these areas create a sexual network of partners which in turn contributes to higher rates of pandemic infection in communities along these routes.

Research has attributed the increase of HIV/AIDS along the main trucking routes to truck drivers. Truck drivers' chance of spreading pandemic is high as they spend most of their time away from their families, and have several sexual partners along their routes. Informal mobility within Botswana also plays a major part in the spread of the epidemic as people move in search for better economic opportunities. This is encouraged by rotation of civil servants and the traditional land tenure. They spend more time away from their families and this can lead them to be involved in unprotected sex.

2.5.3.4 Stigma

Stigma is defined as an imaginary fear of societal attitudes and potential discrimination arising from a particular objectionable attribute disease (Letamo, 2003). Nyblade *et al.* (2001) in their study in Botswana and Zambia revealed that stigma against people living with HIV/AIDS prevented them from participating in prevention programmes like counselling and testing programmes to prevent mother-to-child transmission. There are several reasons why people living with the epidemic are stigmatised. According to Letamo (2003), people living HIV/AIDS are stigmatised because their illness is:

i) Associated with distorted conduct;
ii) Perceived as act of an irresponsible person;
iii) Not well understood by the community and is negatively viewed by health and care givers; and
iv) Viewed as contracted via an immorally behaviour.

In his study to examine the factors influencing stigma and discrimination in Botswana, Letamo (2003) found that HIV/AIDS related stigma and discrimination is widespread. The pandemic on the on-set was a disease that was found in homosexuals, injection drug users and commercial sex workers and these groups were already socially marginalised. Therefore, people living with the epidemic are stigmatised irrespective of how they contracted the disease which has resulted in them being harassed, rejected and exposed to violence. A lot of effort and different measures have been put in place to educate the public on the epidemic. Regardless of this, stigma and discrimination are still extremely remarkable obstacles in preventing the pandemic (Letamo, 2003; Molatole and Thaga, 2006).

2.6. Use of GIS in Health Research

Geographic Information Systems (GIS) is largely a computer software that permits the user to stack or layer several fragments of information for a specific geographic region. All kinds of spatially referenced land data can be stored and manipulated in computer based systems. A

GIS is designed to collect, keep, examine and save data in a structured form. GIS is a technology customarily utilised for resource development and management and it might be very helpful working with data that has spatial attributes. For an overview on HIV/AIDS, GIS and the resultant spatial speculation might be greatly suitable. It can be used to evaluate and interpret the HIV/AIDS epidemic in specific regions and gives the capacity to view spatial associations between information layers that may not be apparent when databases and maps are visually compared. Consequently GIS permits you to maintain and compare spatial attribute in datasets (National geographic society, 2011).

In developing countries, (GIS) for health surveillance data are not well established as compared to developed countries (Kandwal *et al.*, 2009). The use of GIS technology in HIV sero-prevalence research is very limited in Africa. This is partly because of the inadequate means of collecting and monitoring data coupled with limited technological resources (Kalipeni and Zulu, 2008).

However despite its limited use in developing countries, GIS plays a crucial role in confirmatory or exploratory analysis and in the investigation of disease patterns to determine the existence of clusters. Its advancement in the past two decades has empowered researchers in geography and other related disciplines to oversee and investigate a considerable measure of spatially referenced information to assess spatial procedures and designs and to show the after effects of such examinations. Answers to inquiries that worry public health and policy makers can, therefore, be replied by the utilization of spatial clustering procedures (Jongsthapongpanth and Bagchi-Sen, 2010). GIS further helps us to find HIV/AIDS prevalence rates at community level and aids with identifying underserved populations. Additionally it assists state funded organizations to effectively dispense scarce prevention resources to appropriate locations, aid in recognizing at-risk populations and furthermore to determine where to concentrate efforts to prevent HIV/AIDS (Kandala *et al.*, 2012).

The success of GIS has been seen in its application in many areas of population health. It has also been successfully applied in the surveillance and monitoring of vector-borne diseases, in quantifying lead hazards in a neighbourhood (Hanchette, 1999), in predicting child pedestrian injuries (LaScala *et al.*, 2004) and in analysing disease policy and planning (Riner *et al.*, 2004). Geographic mapping could give new understanding into communicable-disease. This was exhibited by Snow (1855) in his study that mapped cholera deaths in relation to London's water pumps. The exercise did not only track the source of outbreak of cholera ,but it also convinced authorities to take action against the disease and demonstrated to future epidemiologists the value of maps as both research and a communication tools etiology and intervention, helping disease control efforts (Snow, 1855).

GIS technology was used to ascertain if existing sexually transmitted infections or AIDS prevention programs were suitably placed for intervention in Cape Town, South Africa. The study also used GIS for identifying where individuals with high rates of new partner acquisition could find prevention programs (Weir et al., 2002). In attempting to estimate morbidity and mortality, GIS was used to generate malaria models that show its occurrence according to seasonal patterns and the intensity of the disease transmission (Tanser and Le Sueur, 2002). Tanser *et al.* (2000) used GIS to assess heterogeneity of HIV/AIDS prevalence among pregnant women whose homesteads are close to the road in the Hlabisa health district of South Africa. The study showed that there is a correlation between HIV/AIDS prevalence and proximity to a primary or secondary road.

In East Africa Ferguson and Morris (2007) used GIS to track the transactional sex trade and truckers along the major highways. GIS has been used to model the utilization of health services with a view to increase the effectiveness of malaria treatment coverage in Kenya (Noor *et al.*, 2003). Geographic information systems have also been used to approximate and examine the spatial distribution of more infectious diseases, including TB in the United States (Moonan *et al.*, 2004). Another study which used GIS estimated local-level HIV prevalence rates using data obtained from antenatal care providers, and the results showed that there is association between HIV prevalence and proximity of local households to a primary or secondary road (Tanser *et al.*, 2000). Econometric procedures were applied to approximate the spatial correlation of HIV infection across international boundaries (McCoskey, 2003).

GIS and HIV sentinel data for pregnant women attending ANCs in Zambia were used by Moise and Kalipeni (2010) to analyse spatiotemporal patterns of HIV prevalence from 1994 to 2004. Spatial statistical regression modelling was used on health-related data to identify possible drivers for HIV prevalence for the year 2004 at district level. They identified possible drivers of HIV prevalence for 2004 at the district scale, where health-related data are reported and planning and decision making were done using spatial statistical regression modelling. Most studies on HIV/AIDS epidemic have been seen to concentrate on the biomedical aspect of the disease instead of focusing on how to better approximate the magnitude of the disease geographically (Moise and Kalipeni, 2010).

GIS has been used in many public health associated areas such as infectious disease control, environment health research, planning and operations, nutrition and physical activity, child and youth health, health promotion, chronic disease prevention, alcohol and other drug services, injury, mental health, vector control, and public health service planning (Queensland Health, 2005). In communicable disease control, Glass (2000) combined GIS and epidemiologic techniques to identify and uncover environmental risk factors associated with Lyme disease in Baltimore County, Maryland.

The etiology of asthma prevalence and residential exposure to air pollution was examined using GIS at a major US-Canada border (Oyana *et al.*, 2004). GIS was used to explain the pollution of drinking water using nitrate, arsenic and chloride in Thessaloniki, Northern Greece (Fytianos and Christophoridis, 2004). GIS was utilised to analyse hot-spot areas and potential pollutant sources in an urban community in Hong Kong. (Li *et al.*, 2004). The exact locations of supermarkets in a mid-sized Canadian city in 1961 and 2005 were mapped utilising GIS (Kristian and Jason, 2005).

The United States National Cancer Institute (NCI) utilized GIS to create geographic patterns for cancer information that had been accessible in tabular form for a considerable length of time. The NCI's inventive work contributed essentially to understanding regional contrasts in rates of lung and oral cancers. In this manner, NCI has made broad utilization of GIS for spatial investigation and perception of information and in addition database improvement. In this study, GIS was utilized to compute distances between residences and hazardous waste sites (HWS) containing pesticides. The study found an expanded breast cancer hazard for ladies living inside one mile of HWS containing organochlorin pesticides (O'Leary *et al.,* 2004). Additionally the survival rates for women diagnosed with breast cancer between 1992 and 1994 were examined using GIS.

GIS has some advantages over other conventional methods in handling spatial data, database management and data analysis, hence its adoption in planning, investigating and estimation of HIV/AIDS incidence patterns. Studies have shown that GIS is a powerful tool not only in analysing disease problems, but in a variety of situations from soil contamination, assessment of potential pollutant sources in an urban community to locations of industrial buildings, buildings and landscapes.

2.7. Limitations of Geographic Information System

Despite the many advantages of using GIS, there are also some limitations. The limitations can be grouped into four classes:

2.7.1 Problems to do with the GIS data model

In most cases data will be archival GIS data or historical and this data may not be accurate (Sipe and Dale.2003). They are likely to be numerous errors in the capturing of data and the accuracy of the original source will be lost when the data are captured. For example when a feature is represented by a crude, hand-drawn, thick line on a map its accuracy maybe questioned. In the GIS it will simply appear as a digital line like any other. Less obvious, but at least as important, is the scale of the source map: a map is only ever accurate within the limitations of its scale (Gregory, 2003). Data taken from maps can be zoomed or integrated using different scales in a GIS. However this demand more from the data than the original map or maps were designed to accommodate and may lead to inaccuracy, error and misunderstanding (Sipe and Dale, 2003). Although historians will be familiar with issues associated with the accuracy of transcriptions, GIS is particularly demanding of the accuracy of its source data (Gregory, 2003).

2.7.2 Problems to do with data itself

There are four types of graphic primitive spatial data, namely; points, lines, polygons or pixels (Gregory, 2003). GIS is more effective when data is defined into locations that are realistically and present the feature to be modelled. That is, if data is not represented in one of the graphic primitive form mentioned above, it cannot be represented. For example, GIS is well suited to modelling hospitals and census districts, but is not well suited to representing the catchment areas for the hospitals where these are poorly defined and overlap heavily with surrounding catchments.

2.7.3 Problems with academic paradigm

Its role in academic geography has yet to be fully established, and history trails some way behind this. Through the 1990s there was considerable debate in geography about whether GIS offered a cohesive, scientific framework that could re-unite the subject, or whether it was a return to a naively positivist agenda (Gregory, 2003). From the historian's point of view, GIS offers new tools, new techniques and new approaches. These approaches must be used critically and should complement traditional ideas, approaches and concerns.

2.7.4 Practical problems

GIS software is expensive and maybe difficult to use. Also, its hardware is still expensive and the data is also expensive to buy and use. GIS data are often financially expensive to buy and capturing them yourself is costly in time as well as money. People with GIS training are often very employable and thus expensive. As a result, entering into GIS is often more costly than originally anticipated and should be done with care (Gregory, 2003: Sepi and Dale 2003).

2.8. Spatial Analysis

There is now an increasing body of literature on spatial analysis of health system and outcomes in developing countries (Adebayo, *et al.*, 2004; Balk, *et al.*, 2004; Kandala, 2006; Kazembe and Namangale, 2007; Gemperli *et al.* 2004; Gemperli and Vounatsou, 2003). Spatial analysis is used to detect health related issues and to distinguish clusters of high or low disease prevalence (Waller and Gotway, 2004).

2.8.1. Types of spatial statistics

Spatial statistics are divided into two categories, global and local spatial statistics (Anselin, 1995). Global spatial statistics summarises the spatial autocorrelation over the entire region of the study area, while local spatial statistics are used to find cluster at local level. The most common spatial statistics are global Moran's I, Getis-Ord General G and Geary 'C. They are used as spatial indicators of spatial patterns of diseases by incorporating the location of the case and the number of cases (Moran, 1948; Moise *et al.*, 2015). These statistics help to identify clustered or dispersed patterns in a particular study area.

The most common tests for local spatial autocorrelation are Getis-Ord Gi* (Hot spot analysis), Anselin LISA and the Kulldorff's Scan statistics. These are used to detect disease hotspots by measuring whether disease incidences are geographically clustered and distinguish between clusters of high value and low value. Local spatial statistical test are essential for health studies as they reveal spatial dependency in some localities (Dean et al., 2005: Moise et al., 2015: Jeefoo, 2012). The important purpose of detecting spatial patterns is that geographic size and shape of clusters, locations of spatial outliers and boundary shapes are identified through the use of spatial statistics (Fotheringham, et al., 2002).

2.9. Literature on Spatial Analysis

The ever growing spatial insight and databases from demographic, political, environmental, ecological, topographical, hydrological, climatic, land-use, public infrastructure, transportation, health infrastructure, and epidemiological sectors have greatly impacted the health arena (Boulos, 2004). In Africa, only a few studies have utilized geospatial analytical techniques in the identification of spatial clusters of high incidence or prevalence. The uses of geospatial analysis of disease have been underutilized in investigating the incidence and spread of HIV/AIDS in Botswana.

Ramjee and Wand (2014) investigated three types of STIs among sexually active using SaT-Scan in Durban, South Africa. Their study revealed that STI occurrence and prevalence was grouped in localized locations which intersected with areas of high HIV prevalence. Spatial patterns of HIV disease among women in Kwa-Zulu Natal province, South Africa, were examined by utilization of geo-additive models. These methods identified important spatial patterns that could not be represented for by standard regression procedures (Wand *et al.*, 2014).

A study by Tanser *et al.* (2009) on localized spatial clustering of HIV infections in widely disseminated rural South African epidemic in Kwa-Zulu Natal, using Kulldorff spatial scan statistic revealed that HIV cases are intense in urban township and high density settlements close to the national road. The study demonstrated that high-density settlements in the south-east of the study area close to the National Road have the most noteworthy HIV prevalence (>35%), whilst the unreachable rural areas close to the western boundary have the least HIV prevalence (<10%). Hixson *et al.* (2011) also studied the spatial clustering if HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations using Kulldorff spatial scan statistic. Their study revealed that HIV epidemic is concentrated in one large cluster which is centralized in downtown Atlanta and this cluster is characterized by poverty, men who have sex with other men (MSM) and IV drug use.

Exploratory spatial data analytical (ESDA) techniques revealed a significant cluster of localized HIV/AIDS in a study conducted by Djukpen (2012) to determine variation of HIV/AIDS in Nigeria. The results showed different epicentres for HIV/AIDS prevalence rates and the spatial clustering maps. In Nigeria significant spatial clusters of HIV/AIDS were discovered by the utilization of Global and Local Moran's I, and Getis and Ord G^*. Fulchera and Kaukinen (2005) also used exploratory spatial data analytical (ESDA) techniques for mapping and visualizing the location of HIV service providers in Toronto. Their findings found a significant clustering of some types of HIV-related services (such as emergency and preventive services), while other services such as medical and end-of-life services are more evenly distributed across Toronto. The identification of HIV infection clusters improved the understanding of HIV infection determinants and geographic patterns. This also contributed to improved allocation of public resources in the Democratic Republic of Congo (Messina *et al.*, 2010).

A number of studies have shown that spatial analyses such as cluster analyses, autocorrelation and kriging are important in understanding disease risks, transmission dynamic, control and surveillance. This has been illustrated by Ruiz *et al.* (2002) who studied the West Nile virus in

Chicago, on bivariate mapping of Lyme disease in the United States, Kitron *et al.* (2006) on the spatial epidemiology of schistomiasis in Kenya and Glavanakov *et al.* (2001) on Lyme disease in New York State. Ruiz *et al,* (2004) investigated effects of environment and socio-political factors on the uneven distribution of the 2001 West Nile virus (WNV) outbreak in the Chicago metropolitan area. Their study used logistic regression equations, but clearly dealt with issues of autocorrelation. A local Moran's I measure of spatial autocorrelation, known as the Local Indicator of Spatial Association (LISA), was utilised to differentiate clusters of cases, or "hot spots," from cases that happen outside the focal areas. The results showed two most remarkable areas with elevated concentration of human cases are in the north part of the study area near the cities of Skokie and Evanston and in the south, around Oak Lawn and Evergreen Park.

Moise and Kalipeni (2010) in their study on the spatial investigation of HIV/AIDS prevalence in Zambia used inverse distance weighted interpolation (IDW). IDW was utilized to spatially interpolate HIV surveillance data into continuous smooth maps based on the 4- year district level HIV prevalence data. The study also used Moran's I statistics to identify specific observation or districts that exhibit spatial autocorrelation with their neighbours. The results reveal that urban areas have higher prevalence than rural areas and the Moran's I statistic indicate that districts that share the same border are more similar with respect to HIV prevalence. Messina *et al.* (2010) conducted a study on spatial and socio- behavioural patterns In the Democratic Republic of Congo a study on socio-behavioural and spatial patterns of HIV/AIDS was conducted using Poison- based spatial scan and inverse distance weighting to detect communities with higher HIV prevalence and to visualize the spatial patterns of HIV prevalence.

Chimoyi and Musenge (2014) studied spatial investigation of factors correlated with HIV infection among young individuals in Uganda using Poison - based spatial scan statistics to detect clusters. The authors additionally utilized spatial binomial logistic regression via Bayesian estimation based on Markov chain Monte Carlo (MCMC) simulation to adjust for spatial and non-spatial random effects. Their study revealed that marital status, age at sexual debut, STIs, alcohol use and condom use were significant predictors of HIV/AIDS infection among persons aged 15-24 years in Uganda.

Bayesian geo-additive mixed model based on Markov Chain Monte Carlo methods were used in the study to examine the geography of HIV/AIDS in 26 districts of Botswana. The study revealed that HIV is varied with the highest prevalence recorded in Selebi-Phikwe and North-East (41.6%) followed by Tutume (41.1%), Bobirwa (39.6%) and Chobe (39.3%). has the lowest HIV prevalence rate of 16.1% (Kandala *et al.*, 2012). In another study, Kandala (2006) used Bayesian geo-additive logistic model to examine the spatial clustering of diarrhea and fever morbidity in Malawi. The researchers were able to identify a distinct pattern of childhood morbidity per district for both diarrhea and malaria. Geo-additive Bayesian semi-parametric models were used to investigate the effect of geographic factors and other critical risk factors on diarrhea, cough and fever in Nigeria (Kandala *et al.*, 2007).

The northern and eastern states had elevated prevalence rate of childhood diarrhea, cough and fever while the western and southern states had lower prevalence rates. The study further revealed that children from poor household and those with educated mothers had a lower association with diarrhea. There was lower association with fever for children whose mothers received prenatal care and gave birth in hospitals. (Kandala, 2006). Most researchers have used Moran's Index and G- statistics to assess the presence and nature of HIV/AIDS spatial autocorrelation and to detect the spatial patterns (Hsueh *et al.*, 2012; Zulu *et al.*, 2014; Moise and Kalipeni, 2010; Obidoa and Cromely, 2012).

To analyze the spatial distribution of malaria problem in Ethiopia, Ayele *et al.* (2012 used a generalized linear model with spatial covariates. Their study provided evidence of spatial distribution of geographic risk factors, demographic and socio-economic in malaria incidence. They also observed that it was riskier to live in the Southern Nations Nationalities and People's (SNNP) region than to live in the Amhara and Oromiya regions.

2.10. Spatial-Temporal analysis

There has been considerable previous academic research on spatial distribution of HIV using different methods in different parts of the world, but not so in Botswana. Geographical research on the spatial patterns of the HIV/ AIDS disease have been focused mainly on identifying its geographical pattern, spatial-temporal inclination and its agents of diffusion.

This section presents studies that looked at spatial-temporal, diffusion spread of HIV/AIDS and geographical pattern using different methods.

Studies carried out by researchers on the spatial- temporal spread of HIV/AIDS indicate that HIV/AIDS diffuses hierarchically and contagiously. Obidoa and Comely (2012), who worked on the spatial-temporal diffusion of HIV in Kenya from 1986-1993 point out that the epidemic spread hierarchically from Nairobi, the capital city of Kenya, to Mombasa the second largest city and then to Kisumu. They noted that the epidemic subsequently spreads both hierarchically and contagiously as it progressed. Hierarchical diffusion is the spread of a thought from people or nodes of authority or power to different people or place. The pattern or idea starts in a major city, then the other cities follow a while after. This kind of diffusion usually begins in a populated area and gradually diffuses to less and less populated areas. Contagious diffusion is the rapid, widespread diffusion of a characteristic throughout the population. It begins with one individual or place and it continuous spreading until it is worldwide. Ideas that spread through infectious diffusion start in a populated area, then other cities adopt the idea and they keep spreading from there.

Gould (1993) (cited in Obidoa and Comely, 2012) reports that local diffusion follows a linear pattern in which transportation routes and infrastructures aiding population morbidity play very crucial role in the spread of the diseases. Air, roads, rail and water transport present varying risks for the spread of HIV/AIDS from one geographic location to another. The movement of large numbers of people along transport networks promotes the spread of HIV/AIDS from high intensity areas to low HIV/AIDS intensity areas. A study conducted in Malawi analysing spatial clustering and the spatial- temporal nature and trends of HIV/AIDS prevalence has identified spatial-temporal trends in HIV prevalence at several scales (national, regional, district and sub-district) using spatial analysis (Zulu *et al.,* 2014).

Lam *et al.* (1996) in their study of spatial-temporal spread of the AIDS epidemic from 1982 - 1990 in four regions of the United States concluded that aids diffusion was hierarchical in the early years and contagious in the later years in Florida. AIDS diffuses from urban areas where a high number of people are at risk of contracting it to small towns or market towns along main highways, then to villages housing people commuting to towns or cities.

The above studies have mainly focused on the spatial-temporal trends of HIV/AIDS and some varying reasons for its spread in general. The current study focuses on applying different spatial procedures to investigate the spatial distribution of HIV infection in Botswana and to compare the results from these methods. The study also identifies the risk factors that are associated with unequal distribution of HIV/AIDS for each district using logistic regression. Statistical significant risk factors in the model are tested using the Wald and the log likelihood tests and the odds ratio (OR) was utilized to interpret the importance of the different risk factors.

Ayele *et al.* (2012) studied the spatial distribution of malaria problem in three regions in Ethiopia using generalised linear mixed model with covariance structure. The results of this study gave proof on the spatial distribution of socio-economic, demographic and geographical risk factors in the occurrence of malaria. A study conducted in Uganda among young people to analyse factors associated with HIV infection revealed that high-risk sexual behaviour and HIV/AIDS prevalence are associated (Chimoyi and Musenge, 2014). Additionally the study identified age at sexual debut, condom use, marital status, STI and alcohol as the major predictors of HIV/AIDS infection among persons aged 15-24 years.

A study by Westercamp *et al.* (2010) on the spatial distribution of sexually transmitted infections (STI) and sexual behaviour among the 18-24 year old sexually active men using cluster analysis in Kenya revealed that sexual risk behaviours and STIs are evenly distributed throughout the Kisumu district. However, the study revealed high and low rate geographical cluster of men who used condoms less frequently. There are linkages between disease prevalence and risk injection behaviours as indicated by a study carried out by Kozlov *et al.* (2006) on HIV prevalence among injection drug users (IDU) in St Petersburg. The spatial analysis also identified where resources might be allocated geographically for maximum impact in slowing the HIV epidemic among IDU.

The geographical distribution of HIV-positive patients in Yunnan province was analysed using trend surface analysis and spatial autocorrelation (Peng *et al.,* 2011). The studies revealed that most severe high case load of HIV were in the central-west region of the province. A negative association was revealed in nearby counties in Kunming and counties with similar high-burden were close to each other. A study to explore spatial autocorrelation of HIV/AIDS epidemic in Chongqing using Moran I and Getis-Ord (Gi^*) revealed that the distribution of HIV/AIDS was clustered at the county level with different directional distribution across China from 2003 – 2009 (Zhi-Hang *et al.*, 2011).

Several studies have used inverse distance weighting, kriging and nearest neighbour interpolation techniques to analyse spatial distribution and spatiotemporal patterns. Zulu *et al.* (2014) conducted a study to examine spatial temporal trends in HIV prevalence, to identify and to map the spatial variation/ clustering of factors associated with HIV prevalence at district level for the year 2010. In their study, they used Inverse Distance Weighted (IDW) in Arch GIS to produce smooth surfaces of HIV prevalence for visualization for the (1994, 1996, 1999, 2001, 2003, 2005, 2007 and 2010) and Moran I to test for spatial dependence. Getis-Ord was utilised to identify outliers or the location of clusters and to classify them into either hotspots or cold spots.

Spatial interpolation method uncovered wide spatial variation in HIV prevalence at regional, urban/rural, district and sub-district levels. However, prevalence was spatially levelling out within and across 'sub-epidemics' while reducing significantly after 1999. Joshua *et al.* (2014) in their study on spatial modelling of HIV prevalence among the clients of female sex workers in Tamil Nadu, south India used multivariate logistic regression model to predict the probability for positive cases of HIV among the clients of FSWs. The inverse distance weighting method (IDW) was used in the GIS methodology to aggregate predict HIV probability (prevalence) to cluster/polygon level.

Numerous studies have been conducted on the patterns of diffusion of HIV in Sub-Saharan Africa. Studies have revealed that HIV spreads hierarchically from urban centres to small towns then to villages. The studies have also indicated that communities living near major roads are at high risk of contracting HIV. Studies have identified numerous risk factors associated with HIV infection among women, sexually active men, the failure to use condoms and many more. Very little work on spatial patterns and diffusion of HIV problem in Botswana has been done. The study by Kandala *et al.* (2012) is the only significant study of the geography of HIV/AIDS prevalence rates in Botswana. However the study only analysed the spatial patterns, but did not provide any understanding as to why some have districts have high prevalence and others low prevalence.

The present study differs from precious studies in that it does not only focus on the evaluation of spatial distribution of HIV/AIDS using a single procedure, but it applies several methods (inverse distance weighting (IDW), kriging and natural neighbour interpolation). The study compares these methods to try to find a robust procedure for spatial distribution of HIV/AIDS in Botswana. It also goes on to identify socio-economic, geographical, demographic, and behavioural risk factors associated with the unequal distribution of HIV/AIDS prevalence per district. Furthermore, the findings from this study will help the government of Botswana in making informed decision in policy formulation to avoid a one-size-fits all intervention strategy which may not be viable in a setting with variation in epidemiological and socio-geographical setting.

2.11. Interpolation Methods

Many disciplines have applied spatial interpolation methods and the most famous are from the environment sciences (Zhou *et al.,* 2007; Šiljeg *et al.,* 2015). The performance of spatial interpolation studies have been compared in many comprehensive published studies by many authors (Zhou *et al.,* 2007; Li and Heap, 2008; Heritage *et al.,* 2009; Guarneri and Weih Jr., 2012; Tan and Xiao, 2014; Šiljeg *et al.,* 2015). Among the existing spatial interpolation methods, some studies have shown that geo-statistical techniques perform better (Li and Heap, 2008). There has been a number of studies carried out on the accuracy of spatial interpolation methods especially in digital elevation models (DEMs), but there is still a need to evaluate the performance of these techniques (Zhi-Hang *et al.,* 2011; Šiljeg *et al.,* 2015). According to Li and Heap (2008): Tan and Xiao (2014) and Šiljeg *et al.* (2015), performance of these spatial interpolators need to be investigated as there are consistent findings about their performance.

Several researchers have compared different interpolation techniques (Luo *et al*., 2008; Yao *et al.,* 2013; Wong *et al*., 2004; Chai *et al*., 2011; Borges *et al.,* 2016; Ye *et al.,* 2014; Murphy *et al*., 2009; Yasrebi *et al*., 2009; Azpurua and Ramos, 2010; Mair and Fare, 2011). In their study Yasrebi *et al.* (2009) compared ordinary kriging (OK) and inverse distance weighting (IDW) interpolation methods to predict spatial variability of some chemical parameters. In their study they used powers of 1-5 to interpolate data using OK and IDW and they found that spatial dependence varied within soil parameters. They also compared the results obtained from the two applied interpolation methods using mean error (ME), mean absolute error (MAE) and root mean square error (RMSE). The spatial distribution of soil chemical properties was predicted and mapped most suitably by the ordinary kriging interpolation technique. .

Azpurua and Ramos (2010) compared spline, inverse distance weighting (IDW) and kriging interpolation techniques in the estimation of average electromagnetic field magnitude. Their aim was to determine the method which creates the best representation of reality for measured electric field intensity. Mean square error (MSE) which measures the average of the squares of the errors and what is estimated and mean absolute error (MAE) is quantity used to measure how close predictions are to the eventual outcome were calculated to determine how best interpolated maps represent the observed data were calculated to determine how best interpolated maps represent the observed data. Of the three interpolation methods used, IDW was the one that best estimated the measurement results of the electric field average magnitude. The study used 10% of the total number of measurement (20 points) and for IDW only 2 points were outside the acceptance region while for kriging and spline11 and 15 points were outside the acceptance region. In a study to compare rainfall interpolation methods in a mountainous region of a tropical island, Mair and Fare (2011) used traditional methods and geo-statistical methods in the analysis. They compared Thiessen polygon, IDW and regression with ordinary kriging and simple kriging with varying local means (sklm). The results showed the ordinary kriging outperformed the traditional methods including the simple kriging with varying local means.

Ansari and Kale (2014) discussed different interpolation techniques for crime analysis using GIS. They discussed spatial analysis (kernel density estimation), interpolation methods (inverse distance weighting, kriging, spline and natural neighbour) and mapping cluster (cluster and outlier analysis [Anselin local Moran's I]) and hot spot analysis (Getis-Ord (G_i^*). In their study they described how the concept of mapping, analysis and hot spot detection techniques can improve crime control.

Fourteen spatial interpolation methods were compared using data obtained from a bathymetric survey of Lake Vrana, Croatia (Siljeg *et al.,* 2015). The researchers used deterministic methods (inverse distance weighting, local polynomial function, radial basis function, and completely regularised spline, with tension, multi-quadric function and inverse multi-quadric) and geo-statistical methods (ordinary kriging, simple kriging, universal kriging, disjunctive kriging, ordinary co-kriging, simple co-kriging, universal co-kriging and disjunctive co-kriging). To compare the accuracy of these spatial interpolation methods, they used a method of cross-validation. Several authors suggest the use of this method in order to attain a successful evaluation of accuracy (Oliver and Webster, 1990; Hofierka *et al.,* 2007). The results from their study revealed that the best deterministic interpolation techniques were the multi-quadric method and the best geostatistical method was the ordinary co-kriging method. The current study aimed to investigate the spatial distribution of HIV/AIDS in Botswana using inverse distance weighting (IDW), kriging and natural neighbour (NN) interpolation methods and to determine the factors that influence its spatial patterns. The study compared the three methods using MSE and MAE to determine the interpolation method that will best describe the spatial distribution of HIV/AIDS in Botswana using data from BAIS IV.

2.12. Conclusion

The above literature review looked at the prevalence of HIV/AIDS worldwide, in Africa and in the study area which is Botswana. The chapter examined research on spatial distribution, GIS and interpolation methods. The chapter also looked at the strength and limitations of methods that were used in different researches. GIS has become the most powerful tool in health research both in developed and under developed countries. However, in Africa GIS is limited due to inadequate means of data collection and monitoring. A number of interpolation methods have been used and compared. These include IDW, OK, NN, spline, and local polynomial function among others. OK proved to be the most efficient interpolation technique.

The studies reviewed compared the performance of interpolation methods of digital elevation model, estimating soil moisture, estimating air quality and crime analysis. There have been limited studies that have assessed the performance of interpolation methods using HIV/AIDS data. As indicated in the assumptions, geo-statistical methods are best when data are normally distributed and stationary. In the reviewed studies, none looked at these assumptions and hence the results from these studies may not be accurate. There are many kriging methods that are ordinary, simple, universal, Bayesian, co-kriging, disjunctive etc. Ordinary kriging was chosen by researchers without comparing it with others to asses if it was the best method for their data.

The current study utilised BAIS IV data to evaluate the performance of geo-statistical and deterministic interpolation methods. The study differs from the reviewed studies in that it compared the performance of different kriging methods to find the most suitable method. The study and employed exploratory spatial data analysis to check if data was normally distributed before using geo-statistical methods.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

3.1. Introduction

Research design is discussed in this chapter as derived from literature, as well as the research philosophy or paradigm that guides this study. Also discussed is the data used in this study, methods of conducting autocorrelation, of identifying hotspot/cold spot, interpolation techniques and logistic regression.

3.2. Research design

The research design, methodology and methods used in this study depend on the research questions that have to be answered. Furthermore, the researcher's approach in conducting the research was guided by how she thought about the problem and how it was to be studied such that the results and findings were to be of quality in the discipline. The research questions were addressed following the research onion as proposed by Saunders *et al*. (2009) shown in Figure 3.1

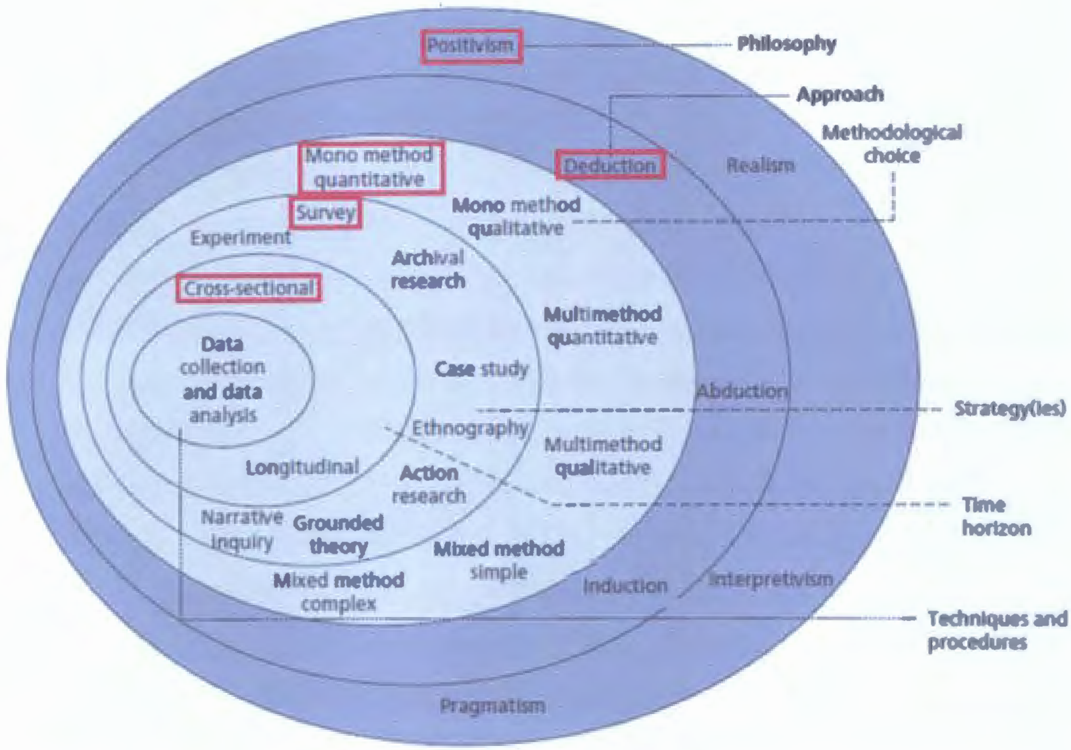


Figure 3.1 Research philosophy in the 'research onion' Source: Saunders *et al.* (2009)

3.3. Nature of methodology

Methodology is the way in which researchers approach problems and seek answers, that is how research is conducted (Taylor, *et al.*, 2015). According to Chilisa and Kawulich (2012), it is where assumptions about the nature of reality and knowledge, values, theory and practice on a given topic get together. Methods are tools used for gathering data and are an important component of methodology (Saunders *et al.*, 2009). This section describes the rationale for the application of specific procedures or techniques used to recognize, select, and examine data applied to understating the research problem, thereby allowing the reader to critically evaluate a study's overall validity and reliability (Labaree, 2009). Methodology is therefore a summary of the research paper that answers two main questions: How was the data collected or generated? And, how was it analysed? (Chilisa and Kawulich; 2012, Labaree, 2009).

3.4. Philosophical Perspective

A good understanding and interpretation of a research philosophy has an exact impact on the research methodology to be adopted in a research study (University of Bradford School of Management, 2003). Some scholars view a paradigm as a model or frame of reference that shapes people's observations and understandings (Rubin and Babbie, 2016). It is a way of examining social phenomenon from which particular understandings of these phenomena can be gained and explanations attempted (Saunders *et al.,* 2009).

A research paradigm guides a research study by outlining the researcher's belief and set of ideas about investigating a research problem and establishing new knowledge in that domain. Guba and Lincoln (1994) explain that it helps answer the following questions; the axiological question which asks, "What is the nature of ethics?", the ontological question which also asks, "What is the nature of reality?", the epistemological question which asks," What is the way of learning and the relationship between the knower and the would be known and finally the methodological which finally asks, "How can the knower go about obtaining the desired knowledge and understanding?" Hart (2010) echoes the same sentiments and reflects that the research paradigm specification guide on"… ontology or a belief in the nature of reality… ways of being, what you believe is real in the world… epistemology, which is how you think about that reality…. methodology, how you are going to use your ways of thinking(our epistemology) to gain more knowledge about your reality and finally, a paradigm includes axiology, which is a set of morals or a set of ethics.

Epistemologically the researcher used her professional knowledge and experience to make sound judgement on the research proceedings and was guided by the ethical values in the arena of research. Guided with this same philosophical approach in claiming new knowledge about a phenomenon, the study developed a comprehensive research design

The research design, methodology and methods used in this study depend on the research questions that have to be answered. Furthermore, the researcher's approach in conducting the research was guided by literature review so as to come out with results that were quality in the discipline. The interrelationship between ontology, epistemology, methodology methods and data (shown on Figure 3.2 below) provides a sound foundation for researchers' guidance.

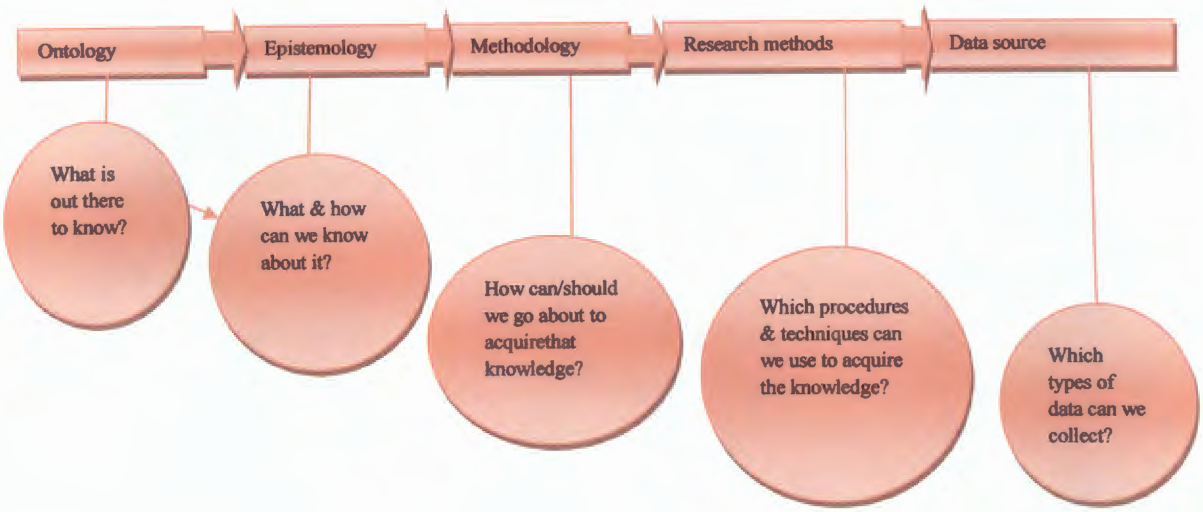


Figure 3.2: The interrelationship between ontology, epistemology, methodology, methods and data Source: Grix (2002:180)

In the current study, the researcher pursued the above outlined logical process by Grix (2002) in which the ontological question was premised on the desire to establish the spatial distribution of HIV/AIDS in selected districts of Botswana. The epistemological orientation focused on the need to determine factors that contribute to spatial distribution. Secondary data was obtained from Central Statistics Botswana.

3.5. Philosophical Approaches in Social Sciences

In social science, there are several predominant paradigms. In this study four most common social scientific paradigms are discussed. These are positivism, social constructionism, critical paradigm and postmodernism.

3.5.1. Positivism

Positivism is directed by the rule of objectivity; know ability, and reasonable logic. Positivism as defined by Rossman and Rallis (2003) as a concept that is directly related with the idea of objectivism. It is a philosophy that uses mathematical data and it includes hypothesis testing to achieve object results (Levin, 1988). Positivism believes that reality is steady and can be illustrated and observed from an objective point without meddling with the phenomenon under study.

3.5.2. Social constructionism

According to Blackstone, 2012:16 "Social constructionist point of view is the possibility that social setting and collaboration outline our realities. Researchers working inside this structure appreciate how individuals come to socially concur or deviate about what is genuine and true". In this paradigm people create reality themselves through their interactions and their interpretations (Blackstone, 2012).

3.5.3. Critical paradigm

In the main, critical paradigm is centred on power, imbalance and social change. "As opposed to the positivism paradigm, the critical paradigm advances that sociology can never be truly objective or value-free. Furthermore, social change should be put into consideration when conducting scientifically investigations" (Blackstone, 2012:17).

3.5.4. Postmodernism

Postmodernism paradigm challenges most social scientific methods for knowing, arguing that there were no universals. The postmodernist paradigm postures a significant challenge for social scientific researchers. While an objective and understandable truth, is guaranteed by the positivism, postmodernists would say that there is definitely not. While social constructionists may contend that truth is subjective depending on each person's preferences (or in the eye of the group that concedes on it), postmodernists may claim that the truth can never be known in a study, this is because the researcher is likely to input his/her own truth on the investigation. At last, while the critical paradigm may argue that reality and truth can be shaped by power, inequality and change, a postmodernist may in turn ask whose power, whose inequality, whose change, whose reality, and whose truth (Blackstone, 2012). Table 3.1 gives a summary of the paradigms and their assumptions.

Table 3.1 Philosophical Approaches

Paradigm	Emphasis	Assumption
Positivism	Objectivity, know ability, and deductive logic	Empirical and scientifically studies can be used to study society.
Social constructionism	Truth as varying, socially constructed, and ever-changing	Our realities can be framed by social context and interaction
Critical	Power, inequality, and social change	Social science can never be truly value-free and should be conducted with the express goal of social change in mind.
Postmodernism	Inherent problems with previous paradigms	Truth in any form may or may not be knowable

Adapted from Blackstone (2012:17)

The researcher in this study has knowledge of statistics and needs to gain more knowledge by objectively analysing and explaining data. The researcher believes that research can be done objectively with no provisions of human experiences. Guided by this ontological belief, the researcher was able to select appropriate statistical and spatial methods to examine the spatial variation of HIV/AIDS in selected districts in Botswana. The researcher has a belief that empirical observations through human sense are very objective and also believes in inductive reasoning. The researcher used data from Botswana Aids Impact Survey IV which is secondary data to attempt explain the concentration of HIV/AIDS in selected districts of Botswana.

3.5.5. Deductive approaches

Deductive approach outlines the steps taken by the research to conduct research. According to Ikeda (2009:53) "deductive process involves the testing of the theory, i.e., it begins with a theory or generalization and seeks to verify whether the theory is applicable to specific cases." Thus, general inferences are used to deduct specific cases. In this approach more specific hypothesis is obtained by narrowing general inferences. These specific hypotheses can be further narrowed down to data collection and hypothesis is done to confirm or deny the theory (William and Donnelly, 2001). In this study the researcher analysed secondary data from 2013

BAIS IV. The aim was to determine the spatial distribution of HIV/AIDS using different methods. This approach enabled the researcher to identify districts with high/low concentration of HIV/AIDS and to formulate recommendation for HIV/AIDS intervention.

Table 3.2 Deductive Research

Theory ⟶ Hypothesis ⟶ Observation ⟶ Confirmation

Adapted from Blackstone (2012)

3.5.6. Quantitative approaches

Quantitative approach emphasises measurement and quantification of data as it mostly applies a deductive approach to the testing of theories (Bryman and Bell, 2015; Saunders and Bezzina, 2015). Regarding the deductive approach, Saunders *et al.* (2009:146) mentioned that:

a) There is the search to explain causal relationship between variables

b) To testing hypothesis, the researcher utilises another characteristic, the collection of quantitative data

c) There should be controls to allow the testing of hypotheses with a control group

d) Research would use highly structured methodology to facilitate replication

e) To pursue the principle of scientific rigour, deduction dictates that the researcher should be independent of what is being observed

f) Concepts need to be operationalized in a way that enables facts to be measured quantitatively

g) Principle of reductionism is followed, that is problem as a whole is better understood if it reduced to the simplest possible elements

h) In order to be able to generalise statistically about regulations in human social behaviour, it is necessary to select sample of sufficient numerical size.

3.6. Study Area

According to Botswana National Atlas, Botswana is a landlocked country in Southern Africa. The country has a maximum length from north to south of about 600 miles (965 km) and a maximum width from east to west of about the same. The country lies between latitudes 17^0 and 27^0 S and longitude 20^0 and 30^0 E. It is bordered on the southeast and south by South Africa, on the west and northwest by Namibia, on the north by Zambia, and on the northeast and east by Zimbabwe. It covers an area of 581.730 square kilometres (Botswana National Atlas). Botswana has a population of 2 024 787 with a population growth rate of 1.26%. The overall life expectancy of the entire population is 54.66 years, with males having a life of 55.75 years and females having a life expectancy of 52.32 years (World Bank, 2012). This country has a birth rate of 21.34 births per 1000 population, a death rate of 13.32 deaths per 1000 population. Figure 3.3 below is Botswana's Political map showing the international boundary, main districts and town councils boundaries with their capitals and national capital.

Figure 3.3 Botswana's Political map. Source: www.mapsofworld.com-2013

3.7. Data

This study used secondary data from the Botswana AIDS Impact Survey IV (BAIS IV), a nationally representative sample survey. The survey was conducted between January and April 2013 by the Central Statistics Office (CSO) in collaboration with the National AIDS Coordinating Agency (NACA) and the Ministry of Health. The major objectives of the survey were:

- To give current HIV occurrence and prevalence estimates between the population aged six weeks to 64 years.
- To give suggestive trends in preventive behaviour among the population aged 10 to 64 years (BAIS IV 2013).

Portable gadgets i.e. smart phones were utilised for the first time information collection in BAIS IV survey. The data collected was immediately entered into smart phones during the field enumeration period and sent to storage centre through network systems, which enabled data capture directly to data servers.

The data that was used in the current study was obtained from Statistics Botswana and was captured in Statistical Package for the Social Sciences (SPSS). The data consist of HIV/AIDS prevalence rates, socio-economic, demographic and behavioural factors for 26 districts of Botswana. Data from SPSS was transformed by converting the file into a database format and then transferred to Geo-Da program. Shape files that were used in this study were also obtained from Statistics Botswana. These files consist of 28 Census districts for Botswana and 26 of the census districts' boundaries. The data base together with the shape files was uploaded into Geo-Da and ArcMap 10.1 for analysis. The overall sample size for the BAIS IV survey was 5415 households. The target population were all individuals aged six weeks and above, and all members who reside in the selected private households aged 10-64 years, excluding the institutional dwellings and industrial areas. Non-citizen tourists who were in Botswana on holiday and not working were also not included in the survey.

Figure 3.4 Shape file map showing 28 census districts of Botswana: Source: Statistics Botswana

3.8. Data Collection

The Central Statistical Office (CSO) adopted a two- stage stratified sampling design for BAIS III. The sampling frame was defined and constituted all Enumeration Areas (EAs) found in the three geographical regions in (i) Cities & Towns (ii) Urban Villages and (iii) Rural Districts as defined by the 2001 Population and Housing Census. The sampling frame consisted of 4114 EAs. During the 2001 Census, EAs were framed of manageable size in terms of dwellings / households, so the primary sampling units (PSUs) were EAs. A list of occupied household in the selected EAs served as a sampling frame for that EA and the secondary sampling units (SSU) were occupied households. Being a two- stage design, two frames were required for each stage. The sampling frame for the first stage was based on the 2001 Population and Housing Census. The sampling frame for the second stage was produced only from the selected EAs by listing of all private habitable dwellings/ households in their EAs. Thus the number of occupied households in the selected EA served as a sampling frame for the EA.

The sampling frame was the 2011 Population and Housing Census and it consisted of listing all Enumeration Areas (EAs) found in three geographical regions which are cities and towns, urban villages and rural districts. The sampling frame was the listing of all occupied households in the selected EA. Stratification was attempted with the end goal that all districts and major urban centres became their own strata.

A two-stage stratified probability sampling design was used for the selection of BAIS IV sample. The first-stage was the selection of EAs and these were selected with probability proportional to the measure of size (PPS). The second stage units were systematically selected from the list of occupied households. Overall, 5415 households were drawn systematically from a listing prepared of selected EAs prepared at the beginning of fieldwork.

3.9.Multi- stage sampling design

Let the population be divided into N first stage units (fsu's) which are disjoint. Define U as $U=\{ U_i,\ldots U_N\}$ for $i = 1, \ldots, N$. The i^{th} fsu U_i is composed of M_i second stage units and is defined as $U_i =\{ U_{ij}, \ldots, U_{iM}\}$. The j^{th} ssu of the i^{th} first stage unit U_{ij} is again composed of M_{ij} third stage units, $U_{ij} =\{ U_{ij1}, \ldots, U_{ijM_{ij}}\}$ and so on.

$N_i =$ total number of EA (fsu) in each district

M_I = Number of households in the ith EA

$M = \sum M_i =$ total number of households in the selecteddistrict

n = number of sampled EAs

π_i = Inclusion Probability of the ith EA

$$= nP_i = n\frac{M_i}{M}$$

m_i =number of sampled households from the ith EA

q_{ij} = total number of individuals (household size) for the jth household of the ith EA

y_{ij} =value of the character y for the jth household of the ith EA , j =1,. . . ,M_i i= 1, . . , N

$$Y = \sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} = \sum_{i=1}^{N} y_{i\bullet}$$

$$\hat{Y} = \sum_{i\in s}\frac{\hat{Y}_i}{\pi_i} = \sum_{i\in s}\frac{M_i\bar{y}_i}{\pi_i} \text{ where } \bar{y}_i = \frac{1}{m_i}\sum_{j\in s_i} y_{ij}$$

$$\hat{Y}_i = \sum_{i\in s}\frac{M_i\bar{y}_i}{nM_i}M = \frac{M}{n}\sum_{i\in s}\bar{y}_i$$

$$\hat{Q} = \frac{M}{n}\sum_{j\in s}\bar{q}_i$$

$$\hat{R} = \frac{\hat{Y}}{\hat{Q}} = \frac{\sum_{i\in s}\bar{y}_i}{\sum_{j\in s}\bar{q}_i} = \frac{\sum_{i\in}\frac{y_{i\bullet}}{m_i}}{\sum_{j\in s} q_{i\bullet}/m_i}$$

$y_{i\bullet}$= total number of individuals who possess character y in the sampled households of the ith EA = $\sum_{j\in s} y_{ij}$

Consider a finite population U = (1,…, N) of N units which have been divided into k disjoint sub-population

Let π_i (>0) and π_{ij} ($i \neq j$) be the inclusion probabilities of the ith and jth units of the population. Let s_j be the set of units that belong to the ith class Ω_i and y_j be the value of a certain character y for the jth (j = 1,…, N) units .Consider an unbiased estimator for the population total

$$\hat{Y} = \sum\frac{y_i}{\pi_i} \tag{1}$$

where the π_i is the inclusion probability of the ith unit and

$$\sum_{s\supset j}\frac{p(s)}{\pi_j} = 1 \tag{2}$$

Using Sarndal (1992) notation, an unbiased estimator of Y = $\sum_{j\in\Omega_i} Y_i$, the total of the y-values of the units belonging to Ω_i is given by

$$\hat{Y}_i = \sum_{j \in s} \frac{y_j I_j}{\pi_j} \qquad (3)$$

Where $$I_j(i) = \begin{cases} 1 \text{ if the ith unit belongs to } \Omega_i \\ 0 \quad \text{otherwise} \end{cases}$$

Consent for participation in the BAIS IV was requested for all components of the study including blood specimen for HIV testing. "Dried blood (DBS) for ages over 18 months were screened for HIV antibodies in a parallel testing algorithm using commercial ELISA test kits- Vironostika-HIV Uni-Form II plus O (OroganonTeknika, Boxtel, The Netherlands) and Murex (Abbott, Wiesbaden, Germany) as per Botswana National Policy on HIV testing. Any specimen that was reactive on parallel ELISA testing was considered HIV antibody positive and any specimen that was not reactive was considered HIV antibody negative. Dry blood specimen for those less than 18 months was tested for HIV virus/ antigen using DNA PCR Roche technology. Any specimen that was responsive was rehashed and if receptive twice then it was analysed as HIV positive. Tests that were non-receptive were analysed as HIV negative. Samples that were regarded as not suitable for testing or had inconclusive results were excluded from prevalence and incidence computations."(BAIS IV, 2013:6) Confidentiality and anonymity were strictly adhered to in handling the data for the survey.

3.10. Conceptual Framework

This study investigates the spatial distribution of HIV/AIDS prevalence in Botswana using BAIS IV data collected by Statistics Botswana in relation to demographic, socio-economic and behavioural factors. These factors form the conceptual framework of the study. A modified conceptual framework is put together using proximate determinants framework by (Boerma and Weir, 2005: Barnighausen and Tanser, 2009 and Lopman *et al.* 2008). Underlying determinants in the conceptual framework act through the proximate determinants and eventually through the biological determinants (Lopman *et al.*, 2008). The direct link between the proximate determinants and the biological determinants influences the infection rate of HIV/AIDS. It also determines the prevalence rate of the disease. The underlying determinants, proximate determinants and the biological determinants have a great influence on the spatial distribution of HIV/AIDS.

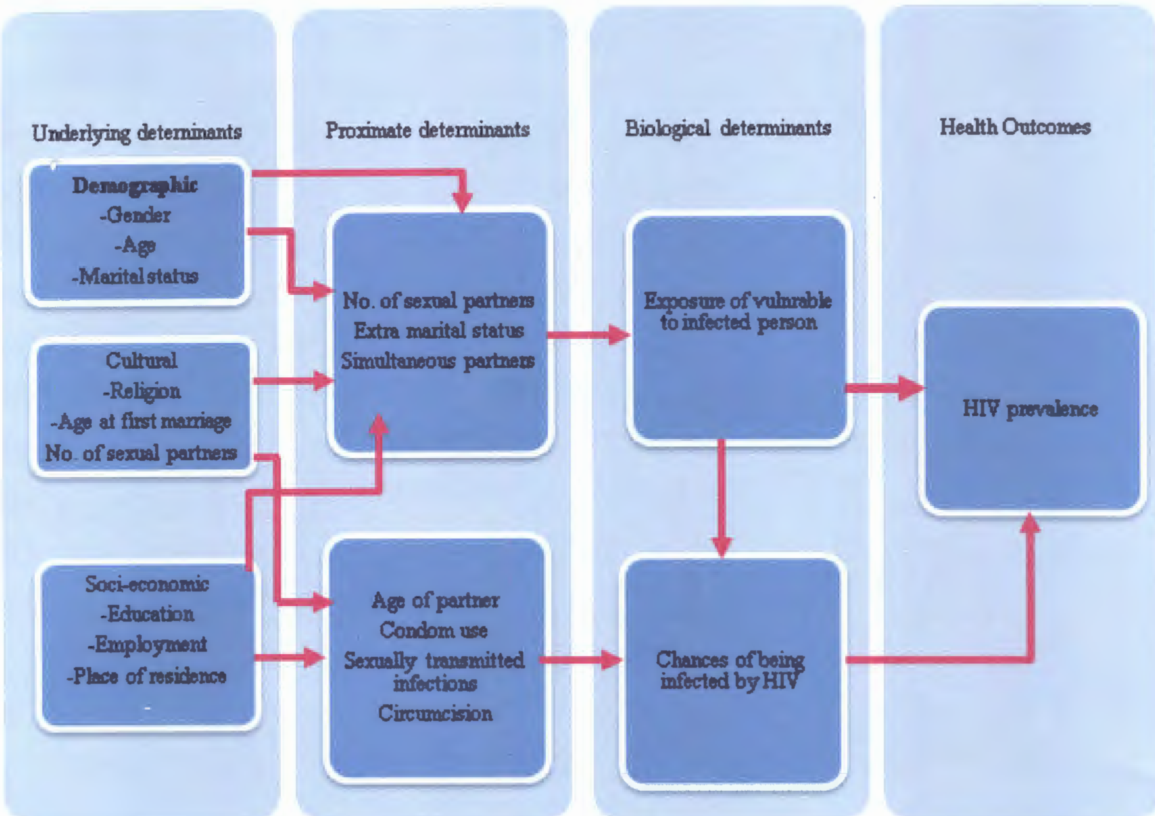


Figure 3.5 Conceptual frameworks for factors contributing to the spread of HIV/AIDS

3.11. Methods of data analysis

Methods that are used to analyse data are discussed. These are quantitative descriptive approach, Moran's I (used to test for autocorrelation) and spatial interpolation methods (inverse distance weighting, ordinary kriging and nearest neighbour) used for predicting and creating continuous surface maps. Also discussed are methods of identifying hotspot/ low spots of HIV/AIDS and logistic regression which were used to determine factors that contribute to the spatial distribution of HIV/AIDS in Botswana.

3.10.1 Preliminary data analysis

Preliminary data analysis is applied to raw data to prepare it for primary analysis. The objective of the analysis is to edit the data to prepare it for further analysis. The key features of the data are described and reliability and validity checks are performed.

3.10.2 Descriptive statistics

Burns and Grove (2005) describe quantitative approach as an objective process that deals with numbers. It is based on traditional scientific methods which generate numerical data. The approach is used to describe variables, examine relationships and to determine cause and effect interactions between variables (Burns and Grove, 2005). Quantitative approach is classified into three groups, descriptive, experimental and comparative approach. Descriptive approach is a basic research method and it deals with describing the characteristic of data using numbers. In this study, preliminary analysis was done using the descriptive approach to describe the overall characteristic of the study population. This approach was further used to compare the relationship between selected socio-economic factors and prevalence in different districts.

Measures of central tendency (mean, mode and median) were used to describe continuous variables; while categorical variables were described using frequencies and percentages. Bivariate analysis which involves the analysis of two variables was used to establish the association between them. For this study, the chi-square test will be employed to determine if there is a relationship between the outcome variable and the explanatory variable. A p-value less than 0.05 will indicate a significant relationship between the outcome and explanatory variables. The analysis was done using SPSS.

3.10.3 Test for reliability

Since cross sectional data was used, there was a need to test for reliability and consistency in the data. Internal consistency is the degree to which the measurement of a test stays uniform over repeated tests of the same subject under similar conditions. An experiment is said to be reliable if the results of the same measure are similar and not reliable if results are different in repeated measurement (Lwanga *et al.*, 1999). To test for consistency and reliability Cronbach's alpha was used. A Cronbach's alpha of 0.6 is reliable and 0.8 or higher indicates good reliability. Very high reliability (0.95) is not desirable as it indicates that an item is redundant. A low Cronbach's alpha suggests that data is not reliable.

There are several reasons why a Cronbach's alpha has a low value. Reverse coding and multiple factors are some of these reasons. In reverse coding two questions may be asking the same thing, with reverse wording. If one uses Likert scale of 1 to 6 with 1 meaning disagree and 6 strongly agree. Suppose the two questions are: I like statistics and I dislike statistics with a response of 1 and 3 respectively. In order to get a reliable Cronbach's alpha, scores of negatively phrased questions need to be reversed. The question I dislike statistics will be scored as 4 (7 minus 3 the recorded score). For multiple factors, questions asking the same thing or factor. The questionnaire needs to be split into tests according to the factors they are testing and a Cronbach's alpha is conducted for each test

3.10.4 Primary data analysis

Primary data analysis is the application of statistical techniques to data that study questions can answer. Different techniques were used to answer the research questions.

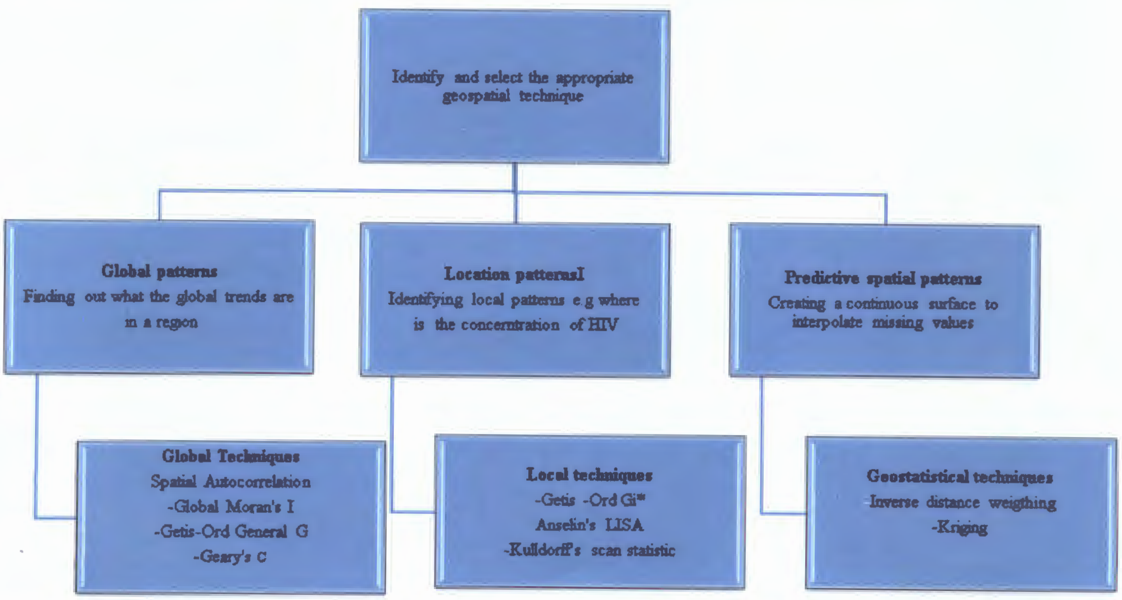


Figure 3.6 Framework for the guide of spatial analysis. Adopted from Geospatial analysis of global health, 2015

3.10.5 Spatial Autocorrelation of HIV/AIDS prevalence in Botswana

Spatial autocorrelation analysis was employed to identify spatial patterns of HIV/AIDS in the districts of Botswana. It was used to test and measure districts that are clustered or dispersed with respect to HIV/AIDS prevalence using BAIS IV. Moran's I statistics and Getis-Ord General G were used to identify spatial patterns of HIV/AIDS for the 12 selected districts of Botswana. These statistics examines the extent of clustering that exists in a dataset observed across the entire region, which is the similarity or dissimilarity of features, based on both their locations and values (Moise et al., 2015).

Moran's I depends on Waldo Tobler's first law of geography which expresses that everything is identified with everything else, close things are more related than far off things (Zulu *et al.*, 2014; Kalipeni and Zulu, 2008). HIV/AIDS prevalence rates should be alike between neighbouring districts than between non-neighbours if the above law applies. Moran's I tests the hypothesis that HIV/AIDS is randomly dispersed versus the alternate that HIV/AIDS is not randomly dispersed. This measure lies between -1 and +1. A Moran's I value greater than zero indicates that HIV/AIDS is positively correlated, a value less than zero indicates dissimilar patterns of HIV/AIDS among adjacent or nearby districts and a value of zero indicates total randomness in HIV/AIDS. According to Islam and Tsuiki (2016) the equation for Moran's I can written as

$$I = \frac{n}{(n-1)S^2 W}\left[\sum_{i=1}^{n}\sum_{j}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})\right] \tag{4}$$

where n is the number of districts,

x_i and x_j are the observations from unit i to unit j about the phenomenon of X

w_{ij} is the i^{th} and j^{th} element of the weighting matrix

$\bar{x}$ is the average of observations from unit i and j

$W = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$ and S^2 is the variance of HIV/AIDS positive participants and can be calculated as,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

The statistical significance test for Moran's I using normal distribution is given as,

$$Z = \frac{I - E(I)}{S_{error}(I)} \tag{5}$$

Where

I is the calculated Moran's I value, E (I) is the expected value and S is the standard error

The expectation is given as:

$$E(I) = \frac{-1}{n-1} \tag{6}$$

3.10.6 Getis-Ord General G statistic

Getis-Ord General G statistic is a tool for determining the level of clustering across the entire study area. General G is used to show distance at which values are clustering at a rate higher than expected by chance (Getis and Ord, 1992; Moise *et al*, 2015). The tool is an inferential statistic and the analysis is explained within the settings of a null hypothesis. The null hypothesis states "there is no spatial clustering of HIV/AIDS in the selected districts". The null hypothesis is rejected when the absolute Z score is big and the p-value is small(< 0.05). A positive Z score means that large values are clustered together in the study area and negative Z score means low values cluster together. In this study Getis-Ord General Statistic was used to determine the extent of clustering in the selected districts of Botswana.

According to Briggs Henan University (2010), the general G statistic of the overall spatial association is given by:

$$G(d) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \forall j \neq i \tag{7}$$

where x_i and x_j are attribute values for the features i and j

$w_{i,j}$ is the spatial weight between the feature i and j

From equation (4) $W = \sum_{i=1}\sum_{j=1} w_{ij}(d)$ for $j \neq i$

$$\text{Then } E[G(d)] = \frac{W}{n(n-1)} \tag{8}$$

According to Getis and Ord (1992) the variance of General G is obtained by considering p! random permutation.

$$E[G(d)]^2 = \frac{1}{(p_1^2 - p_2)n^{(4)}}\left[A_0 p_2^2 + A_1 p_4 + A_2 p_1^2 p_2 + A_3 p_1 p_3 + A_4 p_1^2\right] \tag{9}$$

Where $p_j = \sum_{i=1} x_i^j$ $j = 1, 2, 3, 4$ and $n^{(r)} = n(n-1)(n-2).....(n-r-1)$

The coefficient of A are:

$A_0 = (q^2 - 3n + 3)S_1 - nS_2 + 3W^2$;

$A_1 = -[(n^2 - n)S_1 - 2nS_2 + 3W^2]$;

$A_2 = -[2nS_1 - (n+3)S_2 + 6W^2]$;

$A_3 = 4(n-1)S_1 - 2(n+1)S_2 + 8W^2$;

and $A_4 = S_1 - S_2 + W^2$

$$S_1 = \frac{1}{2}\sum_i \sum_j (w_{ij} + w_{ji})^2 \text{ for } j \neq i$$

$$S_2 = \sum (w_i + w_{.i})^2 \text{ and } w_i = \sum_j w_{ij} \quad j \neq i$$

$$\text{Thus } Var[G(d)] = E[G(d)] - \left[\frac{W}{n(n-1)}\right]^2 \tag{10}$$

3.10.7 Identification of Hotspot

Global statistics are important for detecting spatial clusters that occur in the study area, but they are unable neither to identify where the clusters occur nor to identify differences in spatial patterns within the study area (Moise *et al.*, 2015).To identify clusters and the differences in spatial patterns in smaller areas, techniques for local clustering are used. These are clusters are called hotspots. Hotspot is a condition indicating some type of clustering in a spatial variation (Osei and Duker, 2008). Hotspots are geographical areas where there are excess numbers of HIV/AIDS infection relative to the overall average. It is important to detect hotspot, even if global patterns are not clustered.

Significant clusters of HIV/AIDS in the study area were mapped using two clustering techniques. These are the Getis-Ord G_i^*statistics named after its creators, Arthur Getis and Keith Ord and Kulldorff spatial scan statistics. These techniques have the ability to separate districts of high HIV/AIDS prevalence from districts of low HIV/AIDS prevalence. The spread of an infectious disease can be influenced by clusters that occur randomly (Osei and Duker, 2008). Below, techniques for identifying districts with clusters (hot spots of high or low HIV/AIDS prevalence) are discussed:

3.10.8 Getis-Ord (Gi*) statistics

The local G_i^* statistics is valuable for deciding the spatial reliance of neighbouring observations. The local G_i^*statistic requires the creation of a spatial weight matrix. A spatial weight matrix exhibits the strength of the geographic association among observations in a neighbourhood. Spatial weights can be created based on distance or contiguity. Objects that are points may use distance-based weights, while polygons can utilize both distance and contiguity (Moise, *et al.*, 2015).

In Geo-Da, weight files can be created by Queen Contiguity or Rook Contiguity. Queen Contiguity defines objects as neighbours if they share one or more common point, while Rook Contiguity only defines objects as neighbours if they share a similar border (Moise *et al.*, 2015). In this study, spatial weight were created using Queen Contiguity which always has more neighbours than Rook Contiguity. The statistical significance of Getis-Ord statistics was assessed in Geo-Da by increasing the number of permutations. The output from G_i^* statistic detected spatial clusters of elevated values (hotspots) and low values (cold spots) Bhunia *et al.*, 2013. According to Getis and Ord (1992) Getis-Ord (G_i^*) statistic is calculated as follows:

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d) x_j}{\sum_{j=1}^{n} x_j} \quad j \neq i \qquad (11)$$

x_j the HIV/AIDS prevalence rate for district is j,

$w_{i,j}$ spatial weight between districts i and j,

n is the total number of districts

Implementing standard argument adopted by Getis and Ord (1992) from Cliff and Ord 1973; the value of x_i for the i^{th} point is fixed and (n-1)! Random permutations of the remaining x values of j points are considered. These permutations under the null hypothesis of independence are equally likely.

Now letting X_j be the random variable assigned to point j then

$$P(X_j = x_r) = \frac{1}{n-1}, r \neq i \tag{12}$$

$$E(X_j) = \sum_{r \neq 1} \frac{x_r}{n-1} \tag{13}$$

$$E(G_i^*) = \sum_{j=i} w_{ij}(d) E(X_j) \Big/ \sum_{j \neq i} X_j$$

$$= w_i / n-1 \quad \text{where } w_i = \sum_j w_{ij}(d) \tag{14}$$

Similarly

$$E(G_i^*) = \frac{1}{\left(\sum_j X_j\right)^2} \left[\sum_j w_{ij}^2(d) E(X_j^2) + \sum \sum w_{ik}(d) E(X_j X_k)\right]$$

Since $E(X_j^2) = \sum_{r \neq i} x_r^2 \Big/ n-1$ and $E(X_j X_r) = \sum_{r \neq s \neq i} x_r x_j \Big/ (n-1)(n-2)$

Recall that weights are binary

$$\sum_{j \neq k} \sum w_{ij} w_{ik} = W_i^2 - W_i$$

thus

$$Var(G_i^*) = E(G_i^{*2}) - E^2(G_i^*)$$

$$= \frac{1}{\left(\sum_j x_j\right)^2} \left[W_i(n-1-W_i) j \sum_j x_j^2\right] + \frac{W_i(W_i - 1)}{(n-1)(n-2)} - \frac{W_i^2}{(n-1)^2}$$

Setting $\frac{\sum_j x_j}{n-1} = T_{i1}$ and $\frac{\sum_j x_j^2}{n-1} = T_{i1}^2 = T_{i2}$

then $Var(G_i^*) = \frac{W_i(n-1-W_i)}{(n-1)^2(n-2)} \left[\frac{T_{i2}}{T_{i1}^2}\right]$ (16)

The $Var(G_i^*)$=0 when $W_i = 0$ (meaning there are no neighbours) or when

$W_i = n-1$ (all n-1 observations are within d_j) or when $T_{i2} = 0$ (all n-1 observations are equal) (Getis and Ord, 1992).

3.10.9 Kulldorff's spatial scan statistics

The Kulldorff's scan statistics is a spatial scan statistics technique for identifying and evaluating statistically significant spatial clusters, for example HIV, TB or risk factors for a specific disease (Kulldorff and Nagarwalla, 1995). It can be effectively applied to small or medium sized sets of geographical referenced data (Moise *et al.*, 2015). In this study, Kulldorff spatial scan was used to determine districts that have high clusters of HIV/AIDS prevalence. The scan test is based on a maximum likelihood procedure and it uses a circular window of variable size and location to scan the whole map for areas with high or low rates of HIV/AIDS. The likelihood of observing the number of cases inside and outside each circle is calculated and the most likely cluster will be the one with a maximum likelihood.

The current study will apply the Poisson model devised by Kulldorff (1997). Under this model point are generated by an in homogeneous Poisson process and there is only one zone (Kulldorff, 1997). Let N be a spatial process where N (A) is the random number of points in the set $A \subset G$. As the window moves to the study area it defines a collection of $\mathbb{Z}$ of $Z \subset G$. Under the Poisson model point are generated by an in homogeneous Poisson process and there is only one zone (Kulldorff, 1997).

Thus $N(A) \sim P_o(p\mu(A \cap Z) + q\mu(A \cap \bar{Z}))$ for all A

The hypothesis is given as:

$$\begin{matrix} H_o : p = q \\ H_o : p > q \end{matrix} \quad Z \in \mathbb{Z}$$

Under the null hypothesis, $N(A) \sim P_o(p\mu(A))$ for all A

Now applying the same notations as Kulldorff (1997:185-186) the likelihood for Poisson is given as follows:

$$P(n_G) = \frac{e^{-p\mu(Z) - q(\mu(G) - \mu(Z))} [p\mu(Z) + q(\mu(G) - \mu(Z))]^{n_G}}{n_G!} \quad (17)$$

The density function of a specific location being observed at location x is given as:

$$f(x)=\begin{cases}\dfrac{p\mu(x)}{p\mu(Z)+q(\mu(G)-\mu(Z))} & \text{if } x\in Z\\[2ex] \dfrac{q\mu(x)}{p\mu(Z)+q(\mu(G)-\mu(Z))} & \text{if } x\notin Z\end{cases} \tag{18}$$

Then the likelihood function is given as

$$L(Z,p,q)=\frac{e^{-p\mu(Z)-q(\mu(G)-\mu(Z))}[p\mu(Z)+q(\mu(G)-\mu(Z))]^{n_G}}{n_G!}\times\prod_{x_i\in Z}\frac{p\mu(x)}{p\mu(Z)+q(\mu(G)-\mu(Z))}$$

$$\prod_{x_i\in Z}\frac{q\mu(x)}{p\mu(Z)+q(\mu(G)-\mu(Z))}$$

$$=\frac{e^{-p\mu(Z)-q(\mu(G)-\mu(Z))}p^{n_Z}q^{n_G-n_{Zn}}}{n_G!}$$

The likelihood can also be defined as

$$\lambda=\frac{\text{Sup}_{Z\epsilon zp>q}L(Z,p,q)}{\text{Sup}_{p=q}}=\frac{L(\hat{Z})}{L_o} \tag{19}$$

Thus $L_o=Sup_p\dfrac{e^{-\mu(G)}p^{n_G}}{n_G!}\prod_{x_i}\mu(x_i)=\dfrac{e^{-n_G}}{n_G!}\left(\dfrac{n_G}{\mu(G)}\right)^{n_G}\prod_{x_i}\mu(x_i)$

For the numerator supremum are taken all over p and q for a fixed z and there is maximum when $p=\dfrac{n_Z}{\mu(Z)}$ and $q=\dfrac{(n_G-n_Z)}{\mu(G)-\mu(Z)}$

$$\text{Thus } L_o=\begin{cases}\dfrac{e^{-n_G}}{n_G!}\left(\dfrac{n_Z}{\mu(Z)}\right)\left[\dfrac{n_G-n_z}{\mu(G)-\mu(Z)}\right]^{n_G-n_Z}\prod_{x_i}\mu(x_i) & \text{if } \dfrac{n_Z}{\mu(Z)}>\dfrac{n_G-n_Z}{(\mu(G)-\mu(Z))}\\[2ex] \dfrac{e^{-n_G}}{n_G!}\left(\dfrac{n_G}{\mu_G}\right)^{n_G}\prod_{x_i}\mu(x_i) & \text{otherwise}\end{cases}$$

Then final the likelihood is given as:

$$\lambda=\frac{\text{Sup}_{Z\epsilon Z}L(Z)}{\dfrac{e^{n_G}}{n_G!}\left(\dfrac{n_G}{\mu(G)}\right)^{n_G}\prod_{x_i}\mu(x_i)}$$

$$=\operatorname{Sup}_{Z\in Z}\frac{\left(\frac{n_Z}{\mu(Z)}\right)^{n_Z}\left(\frac{n_G-n_Z}{\mu(G)-\mu(Z)}\right)^{n_a-n_Z}}{\left(\frac{n_G}{\mu(G)}\right)^{n_G}} I\left(\frac{n_Z}{\mu(Z)}>\frac{n_G-n_Z}{\mu(G)-\mu(Z)}\right) \qquad (20)$$

If there is at least one district Z such that $\frac{n_Z}{\mu(Z)}>\frac{n_G-n_Z}{\mu(G)-\mu(Z)}$ and $\lambda = 1$ otherwise $I()$ is the indicator function.

Calculations of the Kulldorff spatial scan statistics were done using Sat Scan version 9.0 where three files were created. These are the case file, population files and the coordinate files. The case file which contains identity numbers (ids) and HIV positive case, the population file which includes the total population under study and the coordinate file which comprises of district names, longitude and latitude locations of clusters.

3.10.10 Spatial Interpolation methods

Interpolation is a procedure used to predict the value of attributes at un-sampled sites using measurement made at point location within the same area. Data from point observations is transformed to continuous fields to enable spatial patterns to be compared. The logic behind spatial interpolation is that usually point nearer to each other in space are more likely to be similar than points further apart. This is also confirmed by Waldo Tobler's first law of geography which expresses that everything is identified with everything else; close things are more related than far off things (Kalipeni and Zulu, 2008).

Several different interpolation methods have been used in the analysis of spatial distribution (inverse distance weighting, nearest neighbour, ordinary kriging, universal kriging, kernel density, voronoi (Thiessen) polygon and many more). These spatial interpolation techniques use numerical models to measure point values of a continuous variable at known areas to forecast values at areas that do not have values, consequently generating a continuous surface (Zulu *et al.*, 2014).

Interpolation techniques fall under two categories: geo-statistical and deterministic. Geo-statistical methods use probabilistic statistical models of the measured point (e.g. kriging), while deterministic techniques (e.g. IDW, nearest neighbour) use the degree of similarity (Kalipeni and Zulu, 2008). Deterministic techniques that compute predictions utilizing the whole data set are called globally and those that utilise measured points inside particular neighbourhoods to capture assumed local variation are called local. In this study, three spatial interpolation techniques were compared and these are: inverse distance weighted (IDW), kriging and natural neighbour (NN). The goal is to determine the method that creates the best continuous surface maps for the spatial distribution of HIV/AIDS in Botswana using the 2013 BAIS IV data. Different methods applied to the same data may produce different surfaces and results, hence the need to compare the suitability of these methods.

The analysis was performed using Geo-Da which are open source soft wares which can be downloaded freely from the internet and ArcMap 10.1. Finally maps that depicting clustering and distribution of HIV/AIDS prevalence were visualised using Geographical Information Systems techniques (GIS). Below is a discussion of the benefits and limitations of the three interpolation techniques that were applied to evaluate the spatial distribution of HIV/AIDS in 26 districts of Botswana. Inverse distance weighting was discussed first followed by kriging and lastly the nearest neighbour interpolation method.

3.10.11 Inverse distance interpolation Method

Inverse distance weighting (IDW) Interpolation estimates the values of unknown points using the distance and values to nearby known points. IDW is a precise technique which enforces the condition that the evaluated value of a point is affected more by close known points than by those more distant away. Measured values located near the prediction location are assigned more weight than farther away points. In IDW measured values located closer to the prediction location are assigned more weight than those further away Weights are inversely proportional to the distance between the observation and the interpolated location (Mair and Fares, 2011). Inverse distance weighting techniques have been used by many researchers for a long time (Kalipeni and Zulu 2008; Moise and Kalipeni 2010; Beck *et al.*, 2002; Webster *et al.*, 1994). A weighting coefficient is used to assign weights to sample points and it in turn controls how the weighting influence will drop down as the distance from the new point increases. During the interpolation process, a greater weighting coefficient results in the point have less effect from the unknown point. As the coefficient expands, the estimation of the obscure point approaches the estimation of the closest observational.[1] According to Wu and Wong (2008), IDW is used to estimate the unknown value of $\hat{y}(R_0)$ in location R_0, given the observed values of y at sampled location R_i and the formula is given as;

$$\hat{y}(R_o) = \sum_{i=1}^{n} \lambda_i y(R_i) \tag{21}$$

where R_0 is a linear combination of the weights, y are the observed values, R_i is the sampled location, n is the total number of points and λ_i are the weights and are defined as;

$$\lambda_i = \frac{d_i^{\ P}}{\sum_{i=1}^{n} d_i^{\ -P}} \ 1 \tag{22}$$

Where; d_i is the inverse distance between R_0 and R_i, p is the power[2]

The power (p) can be any real number greater than 0, but the most practical results are found utilising values from 0.5 to 3. It should be noted that a smaller value of p tends to give estimated average values of R_i and large value of p gives larger weight to the nearest point and smaller weight to point further away (Lu and Wong, 2008).

Hence as $p \to 0$ and $\lambda_i = \frac{1}{n}$ then

$$\begin{aligned} y(R_0) &= \sum_{i=1}^{n} \lambda_i y(R_i) \\ &= \sum_{i=1}^{n} \frac{1}{n} y(R_i) \end{aligned} \tag{23}$$

Now letting R_i to be the nearest neighbour of R_0 and let S_i denote the distance between R_0 and R_i, then $\min\{S_i\} = S_j$. As $p \to \infty$ Lu and Wong (2008) defined the weights as $i = j(S_i = \min\{S_0\})$

$$\lambda_i = \begin{cases} 1 & i = j, (S_0 = \min\{S_i\}) \\ 0 & i \neq j \end{cases} \quad \text{and}$$

$$\hat{y}(R_o) = \sum_{i=1}^{n} \lambda_i y(R_i) = y(S_i) \tag{24}$$

---

[1] QGIS documentation; Spatial analysis interpolation

The default is 2.The degree of local influence also depends on the number of unknown points used in estimation. In IDW, all the predicated values are within the range of maximum and minimum values of the known points (Ansari and Kale 2014). In this study different integer power parameters (p=1, 2, 3, 4, 5) were examined (Yasrebi et al., 2009). The best power value was established by minimising the root mean square prediction error (RMSPE) (Djukpen, 2012). The power value that produces the lowest RMSPE was used in IDW interpolation. A spatial resolution of 1km x 1km was used for the output raster image and 8 surrounding data points for this study. Interpolation was limited to Botswana census district boundaries by creating spatial mask for Botswana (Moise and Kalipeni, 2010).

3.10.12 Kriging Interpolation Method

Kriging is a word utilised by geo-statisticians for a family of generalised least- squares regression methods that use existing data in an identified search neighbourhood to approximate the values at un-sampled locations (Mair and Fare, 2011). Goeff, 2005:1 defines kriging as "the optimal interpolation based on regression against observed z values of surrounding data points, weighted according to spatial weights". It assumes that the distance or direction between sample points reflect a spatial correlation which can be used to explain variation in the surface. Kriging is a multistep process which incorporates exploratory statistical analysis of the data, variogram modelling, generating the surface and optionally exploring a variance surface (Babu, 2016). Kriging is appropriate when distance is spatial correlated or when directional bias of data is known. Kriging is similar to IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location.

In IDW, the weight λ_i depends entirely on the distance of the prediction location. However, with kriging the weights are based not only on the distance between the overall spatial arrangements of the measured points and prediction location but also on the overall spatial arrangement of the measured points (Ansari and Kale 2014). Kriging performs better for large sample because of its probabilistic statistics used in prediction. Reliability of kriging is reduced if the sample is small. Kriging interpolation method uncovers the dependency rules and makes prediction. Ordinary kriging is sometimes called unbiased. Assumptions for using ordinary are 1) Intrinsic or wide sense stationarity of the location $Z(s)$; 2) there are enough observations to estimate the variogram, and 3) the mean of the random location is unknown but constant. Ordinary kriging makes use of models of spatial correlation to calculate a weighted linear combination of observed data resulting in estimates of the observed location or specified unobserved location. The weights are chosen so that the average error for the model is zero and modelled error variance is minimized (Johnston *et al.*, 2001). In this study kriging was used twice first to estimate the spatial autocorrelation of the data and the secondly to make predictions.

To achieve these tasks, kriging involves a two-step process. First, it generates the variograms and covariance function to approximate the statistical reliance of values that depend on the swell when good variograms models are available. The general formula for kriging is given as

$$\hat{Z}(\mathbf{s})-m(\mathbf{s})=\sum_{a=1}^{n(s)}\lambda_a\left[Z(\mathbf{s}_a)-m(\mathbf{s}_a)\right] \tag{25}$$

Where

s, $\mathbf{s}_a$ are prediction locations and one of the neighbouring data is indexed by a

n(a) is the number of data points used to estimate $\hat{Z}(\mathbf{s})$

λ_a is the kriging weight assigned to point z($\mathbf{s}_a$) for estimating location s

m(s), m(s) are expected values of $Z(\mathbf{s}_i)$ and $\hat{Z}(\mathbf{s})$

For Ordinary kriging, the mean is assumed to be constant in the local neighbourhood for each estimation point. That is m(s) = m ($\mathbf{s}_a$). According to Goeff, 2005 ordinary kriging equation is given as:

$$\hat{Z}(\mathbf{s})=m(\mathbf{s})+\sum_{a=1}^{n(s)}\lambda_a\left[Z(\mathbf{s}_a)-m(\mathbf{s})\right] \tag{26}$$

$$=\sum_{a=1}^{n(s)}\lambda_a(\mathbf{s})Z(\mathbf{s}_a)+\left[1-\sum_{a=1}^{n(s)}\lambda_a(\mathbf{s})\right]m(\mathbf{s})$$

Factoring out the mean and letting ordinary kriging weights sum to one gives the following equation for ordinary kriging:

$$\hat{Z}_{ok}(\mathbf{s})=\sum_{a=1}^{n(\mathbf{s})}\lambda_a^{ok}(\mathbf{s})Z(\mathbf{s}_a) \qquad \text{and} \qquad \sum_{a=1}^{n(\mathbf{s})}\lambda_a^{ok}(\mathbf{s})=1$$

(9)

Error variance needs to be minimized subject to the unit sum constraint on the weights and an additional term involving a Lagrange parameter $m_{ok}(\mathbf{s})$ is to be added.

$$L = s_E^2(\mathbf{s}) + 2m_{ok}(\mathbf{s})\left[1 - \sum_{a=1}^{n(\mathbf{s})} \lambda_{\mathbf{a}}(\mathbf{s})\right] \tag{27}$$

Differentiating L with respect to m gives:

$$\frac{1}{2}\frac{\partial L}{\partial m} = 1 - \sum_{a=1}^{n(\mathbf{s})} \lambda_a(\mathbf{s}) = 0$$

Therefore the equations for the kriging weights become

$$\sum_{b=1}^{n(\mathbf{s})} \lambda_b^{ok}(\mathbf{s}) C_R(\mathbf{s_a} - \mathbf{s_b}) + m_{ok}(\mathbf{s}) = C_R(\mathbf{s_a} - \mathbf{s}) \quad \text{a= 1, 2....n}(\mathbf{s}) \tag{28}$$

$$\sum_{b=1}^{n(\mathbf{s})} b^{ok}(\mathbf{s}) = 1$$

where C_R is the covariance function for the residual component of the variable (Goeff,2005).

3.10.13 Variogram

"A variogram is a geostatistical technique which can be used to examine the spatial continuity of a regionalized variable and how this continuity changes as a function of distance and direction. Computation of a variogram involves plotting the relationship between the semi variance *(γ (h))* and the lag distance *(h)."* (Iacozza and Barber, 1999:26; Sunila and Kollo, 2007:2). The best weights for interpolation can be achieved by utilising a variogram (Burrough and McDonnell, 1998; Sunila and Kollo, 2007). Spherical, exponential, linear and Gaussian are the most commonly used variogram models.

Spherical model: One of the most regularly utilised models in geo-statistics is the spherical model (Oliver and Webster, 1990). Optical spherical model can be obtained when the nugget variance is significant but not too large, and when the range and sill are clear (Burrough and McDonnell, 1998).

$$\gamma(h) = \begin{cases} c_c + c_1\left[\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right] & \text{for } 0 < \text{h} < a, \\ c_0 + c_1 & \text{for h} \geq a \end{cases} \tag{29}$$

where $\gamma(h)$ =semi variance, h = lag, a = range, c_0 = nugget variance and $c_0 + c_1$ =sill

Exponential model: This model is a good choice when there is a clear nugget and sill, but only a gradual approach to the range (Sunila and Kollo, 2007).

$$\gamma(h) = c_0 + c_1\left\{1 - \exp\left(-\frac{h}{a}\right)\right\} \tag{30}$$

Linear model: It is a non-transitive variogram as no sill within the area sampled and typical attributes are varying at all scales (Sunila and Kollo, 2007).

$\gamma(h) = c_o + bh$, where b= the slope of the line

Gaussian model: The Gaussian model can be the best model for variogram when the variance is very smooth and the nugget variance is very small compared to the spatially dependent random variation (Sunila and Kollo, 2007).

$$\gamma(h) = c_0 + \gamma(h) = c_0 + c_1\left\{1 - \exp\left(-\frac{h^2}{a^2}\right)\right\} \tag{31}$$

Several types of kriging methods have been developed (simple kriging, ordinary kriging, universal kriging, disjunctive kriging, Bayesian kriging etc.) (Asa *et al.,* 2012). For this study, ordinary kriging simple kriging and universal kriging were compared to find the best kriging method for BAIS IV and the chosen method was compared IDW and Natural neighbour.

3.10.14 Natural Neighbour Interpolation Method

The last interpolation method that was used in this study is the natural neighbour. The technique finds the nearest subset of input samples to an inquiry point and applies weights to them based on equivalent areas in order to interpolate a value (ArcGIS desktop 9.3). This method is also known as Sibson or "area stealing" interpolation. Natural neighbour is as simple to use as nearest neighbour and it provides more precise results. The technique is a weighted moving average which uses geometric relationships in order to choose and weigh nearby points. The points used to assess the value of a characteristic at location x are the natural neighbours of x, and the weight of each neighbour is equivalent to the natural Boissonmat and Cazals (2000) method, it is assumed that p_i is attached a continuously differentiable function h_{pi} from $\Re^d$ to $\Re$ satisfying $h_{pi}(p_i) = 0$. Natural neighbour is defined as:

$$h(x) = \sum_{i}^{n} \lambda^{1+w}(x) h_{pi}(x), \text{ for some arbitrarily small } w>0 \tag{32}$$

Where:

p_i is the natural neighbour

$\lambda_{pi}(x)$ is the continuous function of x

$h_{pi}(x)$ is the hyperplane to a tangent at p_i

The proof by Boissonmat and Cazals (2000) is adopted to prove the proposition that $h(x)$ is continuously differentiable and interpolates h_{pi}

Proof

Since the λ_{pi} and h_{pi} are continuous over $\Re^d$, h is continuous over $\Re^d$, moreover $h(p_i) = h_{pi}(p_i)$ since $\lambda_{pj}(p_i) = 0$ if $i \neq j$ and $\lambda_{pj}(p_i) = 1$

The Cartesian coordinates in $\Re^d$ are denoted by $x^1....x^d$. Since λ_{pi} are continuously differentiable everywhere except at the p_i, h is continuously differentiable at all $x \in P$, we have :

$$\frac{\partial h}{\partial x^j}(x) = \sum \lambda_{pi}{}^{1+w}(x)\frac{\partial h_{pi}}{\partial x^j}(x)$$

$$= 1 + w\sum \lambda_{pi}{}^{w}(x)\frac{\partial h_{p_i}}{\partial x^j}(x)h_{pi}(x) \tag{33}$$

When $x \to p_i$, $\frac{\partial h}{\partial x^j}(x) \to \frac{\partial h_{pi}}{\partial x}(p_i)$ which indicates that h is continuously differentiable.

For every one of the neighbours inside a certain distance from the interpolation location x are considered and the weight of each neighbour is inversely proportional to its distance to x in traditional weighted average interpolation methods, for example inverse distance weighting. There is a certain level of success when utilising interpolation techniques when the data are consistently dispersed. However it is hard to obtain a continuous surface when the distribution of the data is anisotropic or when there is variation in the data density (Ledoux and Gold, 2005). It is difficult to find the appropriate distance to select neighbours when using interpolation methods and requires a priori knowledge of a dataset. Selection of neighbours in natural neighbour interpolation is based on the configuration of the data hence it is not affected by the above issues (Ledoux and Gold, 2005). Figure 3.7 shows the natural neighbour of a given point. Firstly a voronoi is constructed of all given points, and these are represented by light blue coloured polygon. This results in creating a new polygon, beige in colour around the interpolated point (red star). Weights are then created using the proportion of overlap between the new polygon and the initial polygons.

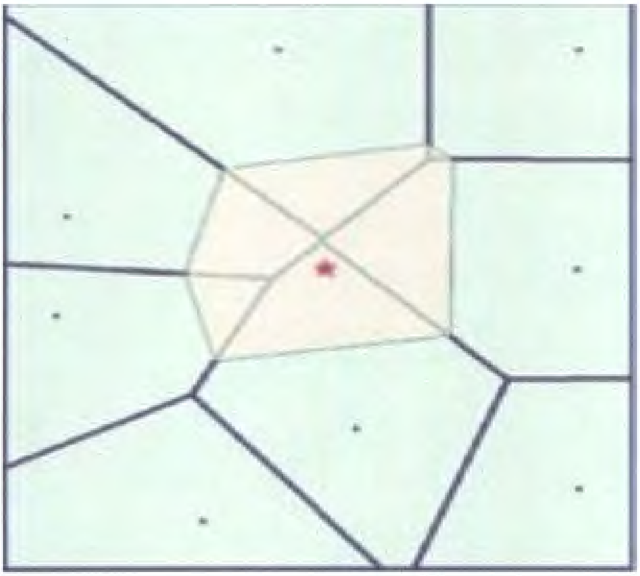


Figure3.7 Natural Neighbour interpolation of a point

3.12. Methods of comparing Interpolation techniques

The aim of the study is to determine the interpolation method that best describes the spatial distribution in Botswana. Interpolated surfaces can be different among various methods and it is not easy to determine the best interpolation method by merely looking at the continuous surface maps (Caruso and Quarta, 1998). As indicated earlier, different spatial interpolation methods applied to the same data set may produce different results, hence the need to assess the performance of these techniques.

Spatial interpolation accuracy can be determined by the method of cross validation. Cross validation is a method for evaluating statistical estimation and prediction (Cressie, 1993). This method is sometimes called the "leave-one-out method since it deletes one sampled point successively and the remaining sampled points are used to predict the missing data points. Usually there is a difference between predicted values and observed values and many errors of predictions can be calculated from the residuals. The following errors were calculated and each interpolation method was compared on the basis of these errors. These are the mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE) and mean error (ME) (Johnston *et al.,* 2001). The best spatial interpolation technique was the one that had small values of RMSE, MSE, MAE and a value of zero for ME if the predictions are unbiased.

The MAE provides an absolute measure of the size of the error and is calculated using equation

$$\text{MAE}=\frac{1}{n}\sum\left|\hat{z}(x_i)-z(x_i)\right| \tag{34}$$

RMSE which provides a measure of the error size that is sensitive to extreme values and is calculated using the equation

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum\left(\hat{z}(x_i)-z(x_i)\right)^2} \tag{35}$$

ME which determines the degree of bias in the estimates and is calculated as follows

$$\text{ME}=\frac{1}{n}\sum_{i-1}^{n}\hat{z}(x_i)-z(x_i) \tag{36}$$

MSE which tests the prediction maps based on measure of accuracy and is calculated as follows

$$\text{MSE}=\frac{1}{n}\sum_{i=1}^{n}\left[z(x_i)-\hat{z}(x_i)\right]^2 \tag{37}$$

where

$z(x_i)$ is the observed value and the

$\hat{z}(x_i)$ is the predicted value and n is the sample size.

3.13. Logistic Regression

Geo- spatial analysis does a good job in determining cluster of HIV\AIDS, but it does not help in answering question why clusters exists in some districts. The analysis cannot answer questions that pertain to identifying factors that contribute to high or low clusters. To help answer such questions, this study used generalised linear model (GLM) approach to determine the socio-economic, demographic and behavioural factors that influence the existence of these clusters. Below is a discussion of the method of generalised linear model.

3.13.1 Generalised Linear Model

The class of GLM are the extensions of the traditional regression models that allow the mean to depend on the explanatory variables through the link function and the response variable to be any member of a set of distributions called the exponential family e.g. Normal, Poisson, Binomial, etc. The GLM approach is attractive according to Jackman (cited by Kurtovic, 2015:4), because it:

- Provides a general theoretical framework for many commonly encountered statistical models.
- Simplifies the implementation of these different models in statistical software, since essentially the same algorithm can be used for estimation inference and assessing model adequacy for all GLMs.

According to Peng and So 2002:33 "Logistic regression (LR) which is suitable for studying the relationship between a categorical variable and one or more predictor variables was used." Logistics regression makes no assumption about the distribution of the independent variables. "The independent variables do not need to be linearly related or normally distributed or have equal variance within the group" (Small, 2009:63). Logistic regression can handle ordinal and nominal data as independent variables. For this study, survey logistic regression which accounts for the complexity of the survey design was used since BIAS IV is a complex survey design. The design involves stratification, clustering and unequal probability of selection. In statistical packages it is necessary to specify strata and cluster before fitting a statistical model. A method of maximum likelihood estimation suggested by Ayele *et al.,* 2012 was used to estimate model parameters.

In logistic regression, the probability of an event happening over the probability of an event not happening is calculated and is called the odds of an event which is given as

$$\text{Odds} \quad (\text{Event}) = \frac{p}{1-p} \qquad (38)$$

Where p is the probability of an event happening and 1 – p is the probability of an event not happening (Peng *et al.,* 2013). "With logistic regression the mean of the response variable p in terms of an explanatory variable x is modelled relating p to x through the equation $p = \alpha + \beta x$ " Park, 2013:155. However extreme values of x give values of $\alpha + \beta x$ which do not lie between 0and 1. This is solved by transforming the odds using the natural logarithms and this is modelled as a linear function of the explanatory variable.

$$\text{Logit}(y) = \ln(\text{odds})$$

$$= \ln\left[\frac{p}{1-p}\right] = \alpha + \beta x \quad (39)$$

where p is the probability of the favourable outcome and x is the explanatory variable. The parameters of the logistic regression are α and β. This is the simple logistic model. Extending the simple logistic regression to multiple predictors gives

$$\text{Logit}(y) = \ln\left[\frac{p}{1-p}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots\dots\dots + \beta_k x_k \quad (40)$$

For the current study the dependent variable was HIV status which was categorised as positive/negative. The predictors were demographic and behavioural factors. Demographic variables included in the study were gender (female/male), marital status (married, never married, separated), age (15-49), education (none, primary, junior secondary, senior secondary, higher education), employment (fulltime, part-time, self-employed, seeking, too old to work, student), knowledge (yes/ no), circumcision (yes/no), religion(Christian, Badimo others).Behavioural factors were age at first marriage (12-49), number of sexual partners (1/>1, condom use(yes, sometime, never) and alcohol use(yes/no). Equation 18 was used to determine the effect of demographic and behavioural risk factors on the HIV/AIDS prevalence rate in different districts in Botswana. Odds ratio was used to interpret the significance of risk factors and 95% confidence interval (CI) to describe the lower and upper limits of the odds. The model was fitted using SPSS 20.

3.13.2 Evaluating goodness of fit

Deviance and likelihood test

The statistical significance of the independent variables (the socio-economic, demographic and geographical variables) was tested. Deviance is calculated by comparing a given model with the saturated model – a model with a theoretically perfect fit (Strikland, 2015). This computation is called the likelihood-ratio test.

$$D_{fitted} = -2\ln L_1 \quad (41)$$

where

$$L_1 = \frac{\text{likelihoodof the fitted model}}{\text{likelihoodof the saturated model}}$$

D represents the deviance

Ln represents the natural logarithm (Strikland, 2015).

The difference between a given model and the saturated model is called a deviance. When the values of the difference are smaller, the model is a better fit (Strickland, 2015). Contribution of a predictor or set of predictors can be assessed by subtracting the model deviance from the null deviance.

$$D_{null} = -2 \ln L_2 \quad (42)$$

where, $L_2 = \frac{\text{likelihoodof the null model}}{\text{likelihoodof the saturated model}}$

Then,

$$\begin{aligned} D_{fitted} - D_{null} &= -2\ln L_1 - (-2\ln L_2)) \\ &= -2(\ln L_1 - \ln L_2) \\ &= -2\ln\left(\frac{L_1}{L_2}\right) \\ &= -2\ln\frac{\text{likelihoodof the fitted model}}{\text{likelihoodof the null model}} \end{aligned} \quad (43)$$

If the deviance model is significantly smaller than the null deviance, then one can conclude that predictors or the set of predictors improve the fitted model significantly.

Wald test

Wald test like the likelihood ratio determine the significance of each predictor variable. It is an approximate of the Likelihood Ratio Test (LRT) with an advantage of testing the joint significance of several coefficients. In this study this test was used to determine whether a certain predictor variable is significant or not.

The Wald test is given by:

$$W = \frac{\hat{\beta}}{Se(\hat{\beta})} \quad (44)$$

Predictor variables that have a large Wald test value and small p-value (0.05) indicate that the variable is significant.

3.14. Ethical consideration

The study uses secondary data of which protocol was submitted to the North-West University Human Research Ethics Committee and permission was granted. Permission was also sought from Statistics Botswana to use the data for analysis.

3.15. Conclusion

In this chapter, the data used in this study and how it was collected were discussed. The study area has been described and various statistical methods used to answer different research questions were presented. These methods include Moran statistic and Getis-Ord statistic to test for autocorrelation cluster respectively. Also discussed are the spatial interpolation methods which include Inverse distance weighting, kriging and nearest neighbour and the software's that will be used. Finally, the logistic regression which was used to identify socio-economic, demographic and behavioural factors that influence spatial distribution has also been discussed.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1. Introduction

The objectives of this study were to ascertain the spatial distribution of HIV/AIDS in some districts of Botswana using different spatial techniques and to determine the best spatial technique for Botswana. The study further intended to determine risk factors associated with the HIV/AIDS in the selected districts and to identify the best spatial technique using 2013 BAIS IV data. The first part of this chapter presents the descriptive and bivariate analysis of the demographic factors followed by autocorrelation analysis which was done using Moran's I and Getis-Ord General G statistics. This is followed by the detection of hot spots using Kulldorff scan, LISA and Getis-Ord Gi* statistics. Interpolation was done using the Inverse distance weighting (IDW), Natural Neighbour (NN) and Ordinary Kriging (OK) interpolation methods and logistic regression were used to determine significant risk factors. The analysis was done using Kulldorff SaTScan, Geo-Da, SPSS version 20 and ArcGIS 10.1 software. Lastly continuous surface maps from the current study are compared with those from previous studies.

4.2. Descriptive Statistics

This section presents the results that help determine if there is a systematic cluster or random spread of HIV/AIDS in some districts in Botswana. To help determine clusters, a quantitative descriptive analysis was carried out; tables and graphs were used to display the results showing the distribution of HIV/AIDS in selected districts of Botswana.

Distribution of HIV/AIDS in selected districts of Botswana

A total of 941098 participants between the ages of 15-49 years were included in the study and of these 116482 were HIV positive. Table 4.1 shows the population density for each selected district and the number of HIV positive individuals. Overall, 12.4% individuals were living with HIV/AIDS in the selected districts. Selebi-Phikwe had the highest percentage (18.6%) of people living with HIV/AIDS followed by Francistown, Central-Mahalapye, Kgatleng and

Central Boteti with prevalence of 15.7%, 13.8%, 13.5% and 13.4% respectively. Central Serowe and Ngamiland South had the least percentage of individuals living with HIV/AIDS with prevalence of 9.7% and 7.5%. Table 4.1 shows that HIV/AIDS prevalence is higher in the city (Francistown) and urban villages (Selebi-Phikwe, Kgatleng, Central Mahalapye and Central Boteti).

Table 4.1 Distribution of HIV positive participants per district

District	Population	HIV Positive	% Positive
Gaborone	127597	15864	12.4
Francistown	67229	10565	15.7
Selebi-Phikwe	32350	6025	18.6
Southeast	47454	5882	12.4
Kweneng East	176158	22217	12.6
Kgatleng	51621	6966	13.5
Central-Serowe	102878	9965	9.7
Central-Mahalapye	85047	11752	13.8
Central-Bobonong	51157	5930	11.6
Central-Boteti	43437	5808	13.4
Central-Tutume	88902	10486	11.8
Ngamiland South	67268	5022	7.5
Total	941098	116482	12,.4

The spatial rates of HIV/AIDS are displayed in Figure 4.1

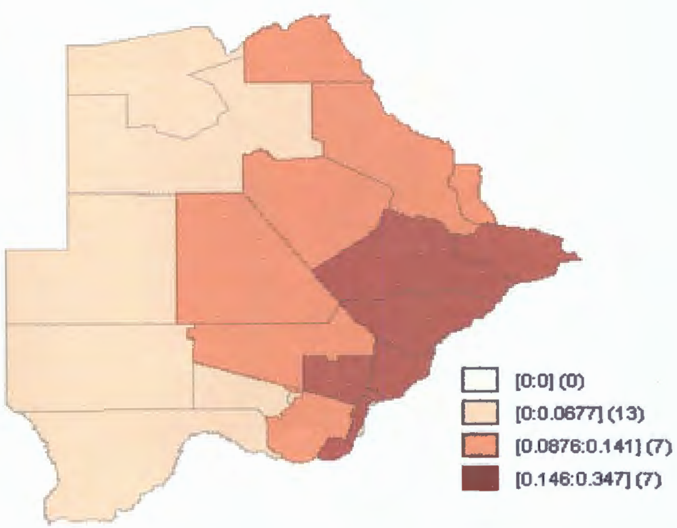


Figure 4.1 showing the spatial distribution of HIV/AIDS in Botswana

Figure 4.1 shows the spatial distribution of HIV/AIDS where a darker brown colour indicates high HIV/AIDS prevalence, while a very light brown colour indicates a lower HIV/AIDS prevalence. Central-Mahalapye, Central-Serowe, Central-Bobonong Kweneng East, Kgatleng, Selebi-Phikwe and Gaborone had the highest HIV/AIDS prevalence followed by Central-Boteti, Central-Tutume and Francistown. Ngamiland south is the only district with least HIV/AIDS prevalence rate in study area. HIV/AIDS prevalence rates are high in the north east and southern part of the country. Botswana is bordered by Zimbabwe in the northeast and east which according to Avert (2016), has the fifth highest HIV/AIDS prevalence (15%) in sub-Saharan Africa. In the south and southeast Botswana is bordered by South Africa which has highest number of people living HIV and a prevalence rate of 12.2% (Avert, 2006; Simbayi *et al.*, 2014).

Distribution of HIV/AIDS in selected Botswana districts by Gender

Table 4.2 HIV/AIDS positive individuals by Gender

Districts	Gender				Total
	Male	Percentage	Female	Percentage	
Gaborone	5951	37.5	9913	62.5	15864
Francistown	3600	34.1	6965	65.9	10565
Selebi-Phikwe	2369	39.3	3656	60.7	6025
Southeast	2733	46.5	3149	53.5	5882
Kweneng East	8470	38.1	13747	61.9	22217
Kgatleng	2657	38.1	4309	61.9	6966
Central- Serowe	4139	41.5	5826	58.5	9965
Central-Mahalapye	4723	40.2	7029	59.8	11752
Central-Bobonong	1836	31.0	4094	69.0	5930
Central-Boteti	1865	32.1	3943	67.9	5808
Central-Tutume	3338	31.8	7148	68.2	10486
Ngamiland South	1823	36.3	3199	63.7	5022
Total	**43504**	**37.3**	**72978**	**62.7**	**116482**

Table 4.2 shows the distribution of individuals living with HIV/AIDS by gender. 62.7% females and 37.3% men were living with HIV/AIDS. Among men living with the pandemic, southeast district had the highest percentage of men living with HIV/AIDS (46.5%) while Central Tutume and Central Bobonong had the least percentages (31.8% and 31.0%)

respectively. Generally more than 50% of females were HIV positive in most districts as compared to men. The highest percentages of women living the pandemic were found in Central-Bobonong (69.0%), Central-Tutume (68.2%) and Francistown (65.9%).

Table 4.3 shows that almost half (47.7%) of the participants living with HIV/AIDS were never married, 37.3% were cohabiting/ living together while 3.8 were either widowed, divorced or separated.

Table 4.3 Distribution of HIV/AIDS individuals by Marital status

	Marital Status								
Districts	Married	%t	Never Married	%	Living Together	Percent	Others	Percent	Total
Gaborone	1388	8.7	5311	33.4	8227	14.7	939	5.9	15865
Francistown	1743	16.5	4175	39.5	4117	39.0	529	5.0	10564
Selebi-Phikwe	906	15.0	2818	46.8	2126	35.3	176	0.9	6026
Southeast	899	15.3	2151	36.6	2703	46.0	129	2.2	5882
Kweneng East	2556	11.5	11265	50.7	8396	37.8	-	-	22217
Kgatleng	468	6.7	4281	61.5	2216	31.8	-	-	6965
Central- Serowe	842	8.4	5358	53.8	3557	35.7	208	2.1	9965
C/Mahalapye	815	6.9	6734	57.3	3681	31.3	521	1.4	11751
C/Bobonong	182	3.1	2636	44.5	2555	43.1	557	9.4	5930
Central-Boteti	682	11.7	2617	45.1	2294	39.5	215	3.7	5808
Central-Tutume	1166	11.1	5248	50.0	2895	27.6	1178	11.3	10487
Ngamiland South	1362	27.1	2926	58.3	734	14.6	0	-	5022
Total	13009	11.2	55520	47.7	43501	37.3	4452	3.8	116482

Ngamiland South, Central Mahalapye, Central Serowe, Kweneng East and Central Tutume having higher percentages of 61.5%, 58.3%, and 57.3%, 53.8%, 50.7% and 50.0% respectively. Among the married participants, 27.1%) live in Ngamiland South 16.5% in Francistown and 15.3% in Southeast Central Bobonong (3.1%) had the least percentage of married individuals living with HIV/AIDS. Gaborone and Ngamiland South had 14.7% and 14.6% respectively, of its participants living together while the remaining districts had over 30% of their participants living together. Central Tutume had 11.3% of its individuals divorced, widowed or separated while Selebi-Phikwe had a least of 0.9%.More than 30% of the individuals were never married in most districts with Kgatleng.

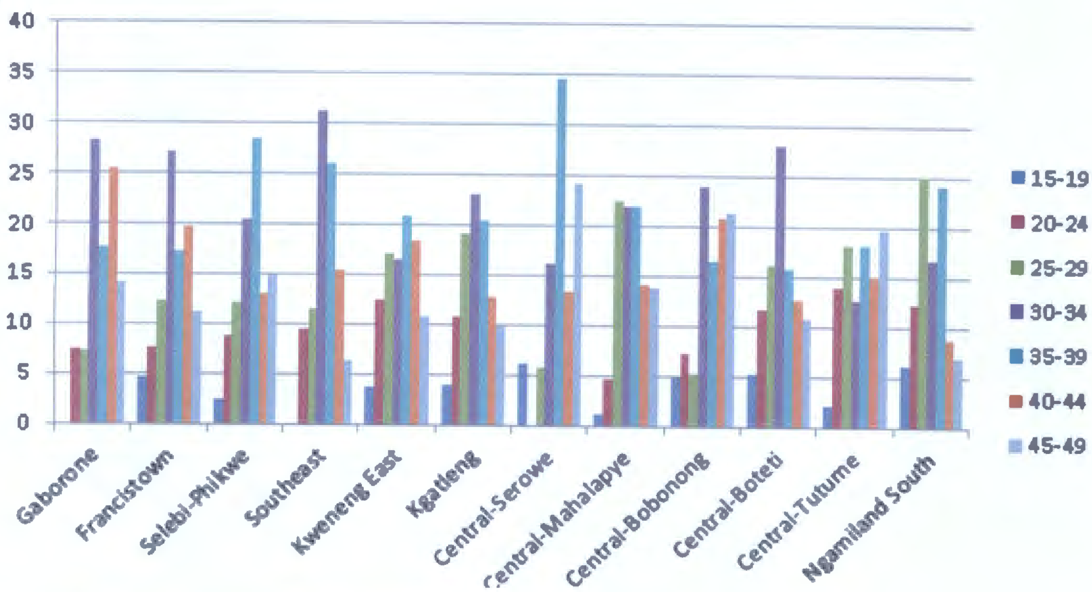


Figure 4.2Distribution of HIV/AIDS in selected Botswana districts by Age

Figure 4.2 it can be seen that HIV prevalence rate is at its peak between the ages 30-34 and 35-39 years with 21.5% and 21.4% The rate is lower for the ages 45-49 years (13.8%).The districts which had higher prevalence rates per ages group were Ngamiland South and Central-Serowe (6.1% and 6.2%) for the age group 15-19 years, Central-Tutume (14.0%) for the age group 20-24 years, Ngamiland South(25.1%) for the age group 25-29 years and Southeast(31.3%) for the age group 30-34 years, Central-Serowe (34.5%) for the age group 35-39 years, Gaborone(25.4%) for the age group 40-44 years and Central Serowe (24.1%) for the age groups 45-49 years. Ngamiland South had lower HIV prevalence rates for the ages 40-44 and 45-49 years and Central Mahalapye for the ages 15-15 and 20-24 years.

Table 4.4 Distribution of HIV/AIDS individuals by Education

District	Level of Education												Total
	NO	%	NON FORMAL	%	PRY	%	JUNR SEC	%	SNR SEC	%	HIGHR	%	
Gaborone	1478	9.3	0	0	3240	20.4	6840	43.1	1056	6.7	3250	20.5	15864
Francistown	612	5.8	0	0	1880	17.8	4422	41.9	2351	22.3	1300	12.3	10565
Selebi-Phikwe	227	3.8	101	1.7	1567	26.0	2210	36.7	1271	21.1	649	10.8	6025
Southeast	410	7.0	315	5.4	430	8.2	2348	39.9	1460	24.8	869	14.8	5882
Kweneng East	607	2.7	0	0	5743	25.8	10944	49.3	1492	6.7	3432	15.4	22218
Kgatleng	0	0	0	0	1560	22.4	3219	46.2	1713	24.6	474	6.8	6963
Central-Serowe	1362	13.7	0	0	2496	25.0	4364	43.8	1102	11.1	641	6.4	9965
Central-Mahalapye	444	3.8	219	1.9	3207	27.3	6667	56.7	490	4.2	724	6.2	11751
Central-Bobonong	450	7.6	0	0	1975	33.3	1862	31.4	1644	27.2	0	0	5931
Central-Boteti	621	10.7	0	0	1127	19.4	3259	56.1	516	8.9	264	4.9	5807
Central-Tutume	773	7.4	168	1.6	3885	37.0	3926	37.4	679	6.5	1050	10.0	10487
Ngamiland South	88	1.8	0	0	1710	34.1	2223	44.3	123	2.4	878	17.5	5022
Total	**7078**	**6.1**	**803**	**0.7**	**28870**	**24.8**	**52284**	**44.9**	**13897**	**11,9**	**13551**	**11.6**	**116483**

Table 4.4 shows that 7078(6.1%) of the participants living with HIV/AIDS pandemic had no education, 803(0.7%) had non-formal education, 28870 (24.8%) had primary education, 52284(44.9%) had junior secondary education, 13897 (11.9%) had senior secondary education and 13551(11.6%) had higher education. This implies that majority of participants living with HIV/AIDS had attained junior secondary school in most districts. In Central Mahalapye and Central Boteti, over 50% of the participants had attained junior secondary school. All most 70% of the participants living with HIV/AIDS in the selected districts had either primary or junior secondary education. All most equal percentages had either senior or higher education. Gaborone, Francsitown, Kweneng East, Kagtleng, Central Serowe, Central Bobonong, Central Boteti and Ngamiland South had no participants with non-formal education. In Kgatleng the least educated participant had primary education.

Table 4.5 Distribution of HIV/AIDS positive individuals by Religion

District	RELIGION						Total
	CHRISTIANITY	%	BADIMO	%	OTHERS	%	
Gaborone	12107	76.3	1906	12.0	1483	9.3	15864
Francistown	9708	91.1	59	0.6	798	7.6	10565
Selebi-Phikwe	5381	89.3	53	0.9	590	9.8	6024
Southeast	4936	83.9	153	2.6	794	13.5	5883
Kweneng East	18903	85.1	1668	7.5	1646	7.4	22217
Kgatleng	4876	70.0	447	6.4	1643	23.6	6966
Central-Serowe	7203	72.3	1009	10.1	1753	17.6	9965
Central-Mahalapye	9116	77.6	444	3.8	2192	18.6	11752
Central-Bobonong	5348	90.2	152	2.6	430	7.3	5930
Central-Boteti	4873	83.9	184	3.2	751	12.9	5808
Central-Tutume	9306	88.7	869	8.3	311	3.0	10486
Ngamiland South	4470	89.0	266	5.3	286	5.7	5022
Total	**96227**	**82.6**	**7210**	**6.1**	**32397**	**27.8**	**116482**

Table 4.5 indicates that HIV/AIDS prevalence rate was high (82.6%) among christians followed by those who had other religion (27.8%) and lastly those of Badimo religion (6.2%). Over 90% of participants living with HIV/AIDS in Francistown and Central-Bobonong districts were Christians. Generally above 70% of the participants were Christians. Gaborone and Central Serowe had the highest percentage of Badimo followers (12.0% and 10.1%) while Francistown and Selebi-Phikwe had the lowest (0.9% and 0.6%). Among participants who belong to other religions, Kgatleng had the highest of 23.6% followed by Central-Mahalapye

and Central Serowe with 18.6% and 17.6%. Only 3.0% of Central Tutume's participants belong to other religions.

Table 4.6 Distribution of HIV/AIDS positive individuals by Employment.

District	FULL-TIME	%	SELF EMPLOYED	%	PART-TIME	%	LANDS/ CATTLE	%	ACTIVELY SEEKING	%	STUDENT	%	PENSENIOR	%	
Gaborone	10722	67.6	1051	6.6	736	4.6	0	0	1177	7.4	368	2.3	1075	6.8	15865
Francistown	4728	44.8	891	8.4	880	8.3	0	0	2631	24.9	609	5.8	573	5.4	10565
Selebi-Phikwe	3292	54.6	467	7.8	335	5.6	120	2	1561	25.9	106	1.8	144	2.4	6025
Southeast	3079	52.3	487	8.3	240	4.1	0	0	1651	28.1	185	3.1	240	4.1	5882
Kweneng East	8050	36.2	2487	11.2	365	1.6	1824	8.2	5497	24.7	1948	8.8	2046	9.2	22217
Kgatleng	3603	51.7	218	3.1	324	4.7	379	5.4	1685	24.2	156	2.2	600	8.6	6965
Central-Serowe	3024	30.3	1041	10.4	991	9.9	1514	15.2	1887	18.9	622	6.2	0	0	9965
Central-Mahalapye	2583	22.0	1005	8.6	494	4.2	796	6.8	5775	49.1	134	1.1	965	8.2	11752
Central-Bobonong	1429	24.1	551	9.3	914	15.4	953	16.1	1505	25.4	182	3.1	396	6.7	5930
Central-Boteti	1124	19.4	811	14.0	399	6.9	557	9.6	2237	38.5	172	3.0	510	8.8	5810
Central-Tutume	3452	32.9	1351	12.9	712	6.8	1808	17.2	1801	17.2	207	2.0	927	8.8	10485
Ngamiland South	2072	41.3	336	6.7	382	7.6	266	5.3	1659	33.0	134	2.7	172	3.4	5021
Total	**47158**	**40.5**	**10696**	**9.2**	**6772**	**5.8**	**8217**	**7.1**	**29066**	**25.0**	**4823**	**4.1**	**7648**	**6.6**	**116482**

Table 4.6 show that HIV/AIDS prevalence was higher among individuals who had full-time employment (40.5%) followed by those who were actively seeking employment (25.0%). For those who were self-employed, worked on part-time, worked in the cattle post, students and pensioners, HIV/AIDS prevalence was below 10% (9.2%, 5.8%, 7.1% 4.1%, 6.6%). 67.6% of the participants living with HIV/AIDS were on full-time employment in Gaborone, 54.6% in Selebi-Phikwe, 52.3% in Southeast, 51.7% in Kgatleng and Central Boteti had the least of 19.4%. Prevalence rate for those who worked on part-time was higher in Central Boteti, Central-Tutume, Kweneng East and Central- Serowe (14.0%, 12.9%, 11.2%, and 10.4%) and Kgatleng had the lowest of 3.1%.

Central Bobonong had 15.4% of its individuals living with HIV/AIDS working on part-time and this is almost three times the overall of individuals living with HIV/AIDS in the selected districts. Prevalence rate was higher in Central-Tutume (17.2%), Central-Bobonong (16.1%), Central-Serowe (15.2%) and least in Selebi-Phikwe (2.0%) for individuals who worked in the cattle post. Generally over 20% of the participants were actively seeking employment with the highest percentages in Central-Mahalapye and Ngamiland South (49.1% and 33.0%). There were more students living with HIV/AIDS in Kweneng East district (8.8%), Central-Serowe (6.2%) and Francistown (5.8%). Central-Mahalapye and Selebi-Phikwe had the least percentages of students living with HIV/AIDS (1.1% and 1.8%). Kweneng East district had the highest percentage (9.2%) of pensioners living with HIV/AIDS followed by Central-Tutume (8.8%), Central-Boteti (8.8%), Kgatleng (8.6%) and Central Mahalapye (8.2%). The lowest percentages of pensioners living with HIV/AIDS were in Selebi-Phikwe (2.4%) and Ngamiland South (3.4%).

4.3. Spatial analysis

In this section first autocorrelation is done to determine spatial patterns of HIV/AIDS followed by hotspot analysis to identify districts with excess numbers of HIV/AIDS relative to the overall average. Finally interpolation is carried out to determine the best method for analysing the spatial distribution of HIV/AIDS using 2013BAIS IV data. The surface maps were created using ArcGIS 10.1.

4.3.1. Exploratory spatial data analysis

There is a need to gain a better understanding of the dataset under study so as to make better decisions especially when creating a surface using interpolation methods. Interpolation methods give best results when the dataset is normally distributed. (Arc GIS 10, 2009).

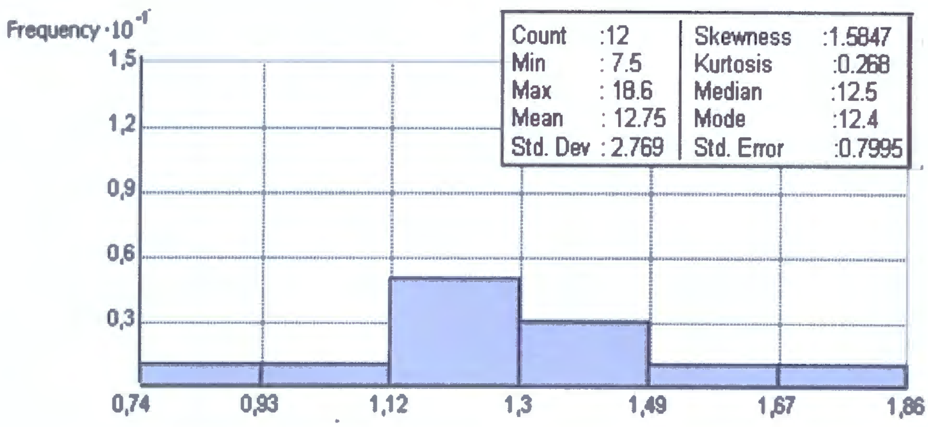


Figure 4.3 Histogram showing distribution of HIV/AIDS

For this study, exploratory spatial data analysis (ESDA) was employed out to examine if the dataset for the study conforms to the above conditions. To test for normality in the data, a histogram was plotted to examine the distribution of the dataset and the results are shown in Figure 4. 3. There is evidence that the dataset is normally distributed since the mode (12.4) almost equal to the median (12.5).

4.3.2. Autocorrelation analysis

The autocorrelation analysis was employed to identify spatial patterns of HIV/AIDS in Botswana selected districts. To help identify this, Moran' I was used to measure the extent of correlation of HIV/AIDS among neighbouring districts and the results are shown in Table 4.8

Table 4 7. Moran's Index for spatial autocorrelation of HIV/AIDS

Moran's I	P-value	expectation	Variance
0.1359	0.0481	-0.0385	0.008

A Moran's Index value of 0.1359 and p-value of 0.0481 for HIV/AIDS in the selected districts of Botswana was obtained. This value is significant and confirms that there were spatial patterns in the occurrence of HIV/AIDS in the selected districts of Botswana. These results were in agreement with the result obtained in descriptive statistics.

4.3.3. Getis-Ord general G

Getis-Ord General G statistic a global statistic that is used to determine spatial clustering by value with respect to the study area was calculated and the results are shown in Table 4.9.

Table 4.8 Getis-Ord General G

Observed General G	z-score	P-value
0.5044	2.4101	0.0160

A z-score value of 2.4101 and p-value of 0.0160 were obtained indicating that the pattern is less likely to be a result of random chance. These results indicate that there was evidence of high cluster patterns in the occurrence of HIV/AIDS in the selected districts of Botswana. Results obtained from descriptive statistics, Moran's I index and Getis-Ord indicated that there were spatial patterns in the occurrence of HIV/AIDS in the districts selected for this study in Botswana.

4.3.4. Hot Spot Analysis results

Hot spot is a condition indicating some form of clustering in a spatial distribution (Osei and Duker, 2008). These are geographical areas that have excess numbers of HIV/AIDS relative to the overall average. In this study hot spots were analysed using Local Indicator of Spatial

Association (LISA), Kulldorff Scan and Getis-Ord Gi* statistics and the results are presented in the next sections.

4.3.5. Local Indicator of Spatial Association results

The LISA cluster and significant maps are shown in Figures 4.4(a) and 4.4(b).

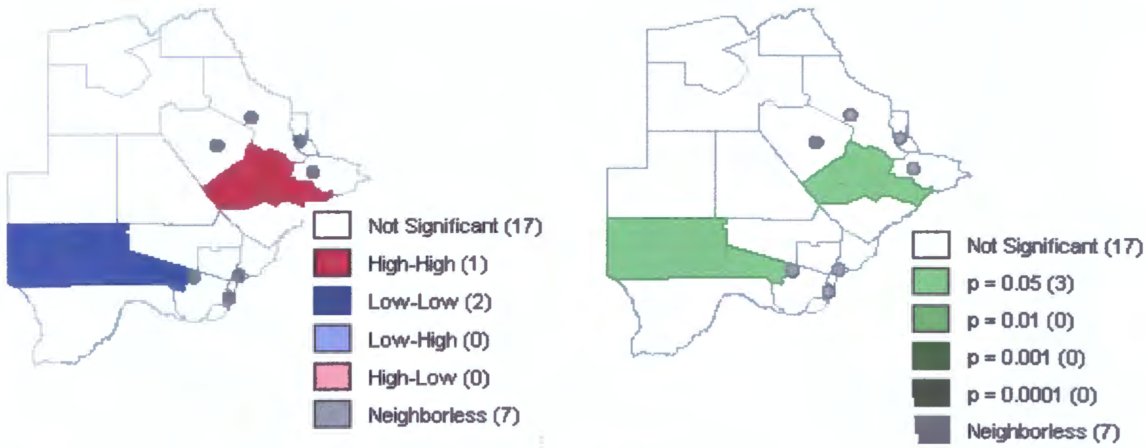


Figure a **Figure b**

Figure 4.4 LISA Cluster Map and LISA Significance Map **(For district names refer to Figure 3.4)**

There is one district with high-high (districts with high HIV/AIDS prevalence rates surrounded by districts with also high HIV/AIDS prevalence rates) and this is Central-Serowe. There are two districts with low-low rates of HIV/AIDS (district with low HIV/AIDS prevalence rate surrounded by districts with low prevalence rate) and these are Ngwaketswe and Kgalagadi north. There are also no districts with low-high (districts with low HIV/AIDS prevalence rate surrounded by districts with high prevalence rate) and high-low (districts with high HIV/AIDS prevalence rate surrounded by districts with low HIV/AIDS prevalence rate). Both clusters are significant at 5% significance level (Figure 4.4b).

4.3.6. Kulldorff's spatial scan results

Kulldorff scan is both deterministic and inferential. That is, it locates clusters and allows hypothesis testing and evaluates the significance. However, its major setback is that it does not allow viewing of created clusters and the centre location and radius for each cluster is available as a text only (Mala and Sengupta, 2013). This is presented in Table 4.9 which is the summary of Kulldorff SaTScan results of HIV/AIDS primary cluster in Botswana.

Table 4.9 HIV/AIDS clusters in Botswana districts using Kulldorff Sat scan

District	Population	No of cases	Expected cases	Annual cases/100000	Observed/Expected	Relative risk	Log likelihood ratio	P-value
Serowe	1273	9963	1411.55	783519.7	7.06	7.63	11248.11	<0. 001
Selebi-Phikwe	769	6025	9234.1	60357.7	0.65	1.36	265.11	<0. 001
Francistown	1350	10565	13670	134692.8	0.77	1.16	100.68	<0.001
Central-Mahalapye	1232.5	11752	7022.81	93716.5	1.67	1.01	0.28	0.998

Figure 4.9 indicates only one primary cluster in the study area with a Log likelihood of 11248.11, a p-value of 0.000000000000001 and a relative risk of 7.63. A risk factor of 7.63 indicated that the risk of getting infected by HIV/AIDS inside the primary cluster (Central-Serowe) was 7.6 times more than the risk outside the cluster. This implies that individuals residing Central-Serowe district were 7.6 times more likely to acquire HIV/AIDS infection than those residing outside the district. In the district there are 9965 number of cases with 1411.55 expected cases and a population of 1273 which is the average population of the cluster which is taken over the whole study period. The annual rate per 100 000 is the number of cases (HIV positive) calculated taking into account leap years and is based on the length of the year.

In addition to the primary cluster, the software detected three secondary clusters and these were Selebi-Phikwe, Francistown and Central-Mahalapye. Individuals residing in Selebi-Phikwe and Francistown were 1.4 and 1.2 times more likely to get infected by HIV/AIDS than individuals living outside the districts. The relative risk for Central-Mahalapye was almost one, indicating that the chance of getting infected inside the district is the same as outside the district. Selebi-Phikwe and Francistown were significant secondary clusters with Log-likelihood 265.11 and 100.68 and p-value less than 0.05 respectively. Central-Mahalapye with a Log-likelihood of 0.28 and a p-value of 0.998 was not a significant secondary cluster.

4.3.7. Getis- Ord Gi* results

Getis- Ord Gi* was applied in ArcGIS 10.1 also to detect hot spot clusters of HIV/AIDS and the results are shown in Figure 4.5.

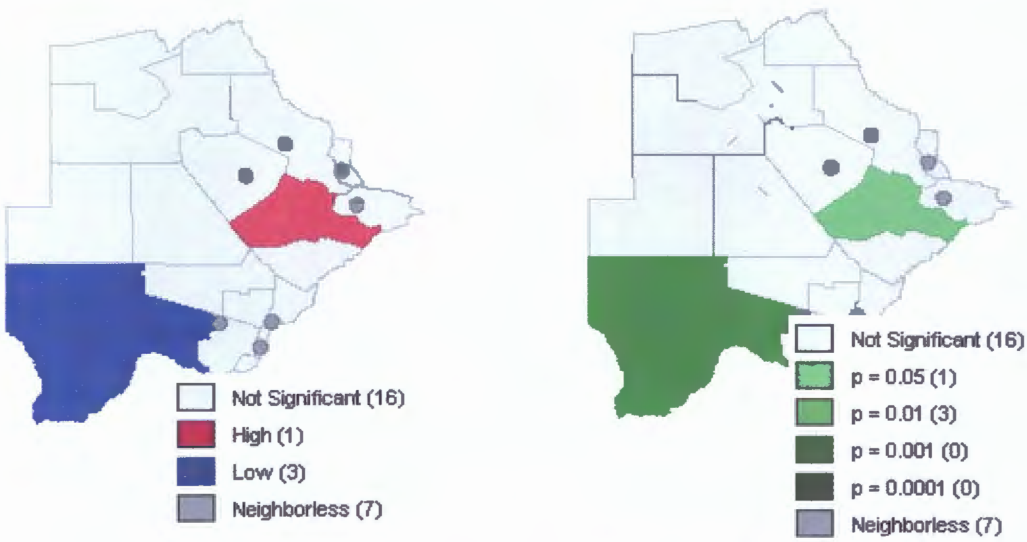


Figure 4.5 Gi* Cluster Map and Significant Map **(For district names refer to Figure 3.4)**

The results show one district that had high HIV/AIDS prevalence rate (Central-Serowe) with a p-value of 0.05 and three district that had low HIV/AIDS prevalence rate with p-value of 0.01. These are Kgalagadi South, Kgalagadi North and Ngwakwetse. These findings are in support of LISA and Kulldorff SaTScan statistics. From these results it can be concluded that Central Serowe was the epicentre of HIV/AIDS in Botswana. An epicentre is an area with high HIV/AIDS concentration. Central-Serowe is surrounded by Central-Mahalapye, Selebi-Phikwe and Francistown. Descriptive analysis showed that these districts have high HIV/AIDS prevalence in the selected study area.

4.4. Interpolation

In this section spatial distribution of HIV/AIDS was performed using three different Kriging interpolation methods and the best Kriging method was chosen. This method together with Inverse Distance Weighting (IDW) and Natural Neighbour (NN) were combined to determine the best interpolation method for analysing the spatial distribution of HIV/AIDS in Botswana using 2013 BAIS IV. The interpolation methods were performed using the geostatistical analyst extension within the geographical information system ArcGIS (Version 10.1.

4.4.1. Spatial distribution of HIV/AIDS using Ordinary, Simple and Universal Kriging methods

Firstly, different kriging methods were used to analyse the spatial distribution of HIV/AIDS with the aim of choosing the best kriging method. Maps showing these kriging interpolation methods are shown in Figure 4.6.

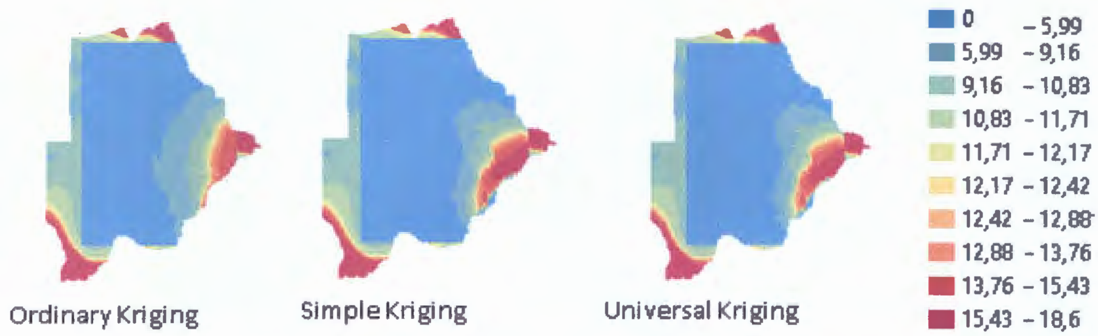


Figure 4.6 Maps of Ordinary, Simple and Universal kriging for HIV/AIDS

Ordinary kriging and universal kriging maps show the same spatial distribution of HIV/AIDS. The maps indicate that HIV/AIDS is concentrated in the southern parts northern and central parts of Botswana. The intensity of the disease becomes less as it spreads out towards the western parts of the country. According to the results displayed in Figure 4.11, it is difficult to choose the best kriging methods. The cross validation procedure which calculates prediction errors was employed to determine the best kriging method.

Table 4.10 shows the estimated errors from the cross validation procedure for ordinary kriging, simple kriging and universal kriging.

Table 4.10 Prediction errors for different kriging methods

Kriging method	RMS	MS	AVE
Ordinary	6.3172	-0.02067	4.1728
Simple	6.3263	-0.0106	4.8147
Universal	6.8089	-0.02295	7.5531

Table 4.10 shows that ordinary kriging had the least values of RMS (6.3172), MS (-0.02067) and AVE (4.1728) followed by simple kriging with RMS (6.3263) and AVE (4.8147). Simple kriging had highest values of RMS (6.8089), AVE (7.5531) and a second least value for MS (-0.02295). Overall, ordinary kriging had the least error values and hence was chosen as the best kriging method to analyse the spatial distribution of HIV/AIDS in Botswana using 2013 BAIS IV. In addition to having the least errors, it is the most basic and popular kriging method (Webster *et al.*, 1994; Zulu *et al.*, 2014 (Sentianto and Triandini, 2013: Azpurua and Ramos, 2010: Gou *et al.*, 2010: Zimmerman *et al.*, 1999).).

4.4.2. Spatial distribution of HIV/AIDS IDW, OK NN interpolation techniques

The interpolation procedure was carried out using the methods: inverse distance weighting (IDW), Ordinary kriging (OK) and Nearest Neighbour. The results of each interpolation technique were represented on the study map in Figure 4.7.

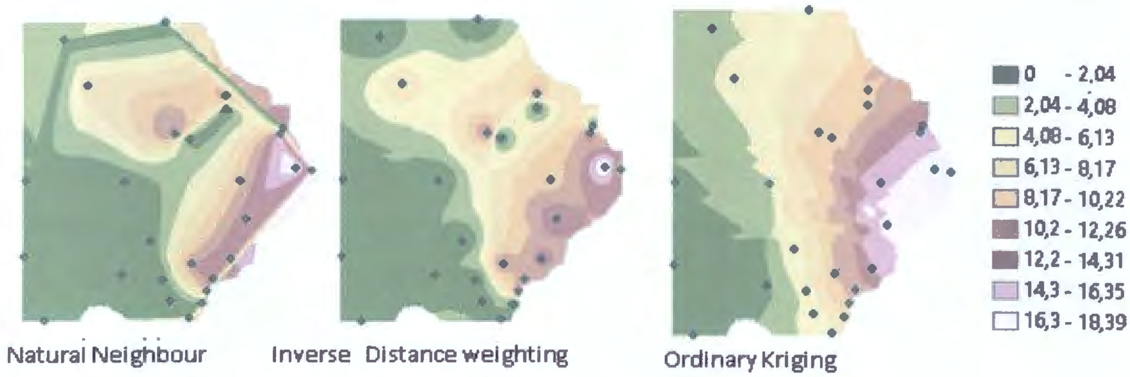


Figure 4.7 Natural Neighbour, Inverse Distance Weighting and Ordinary Kriging Maps

The dots represent the districts of Botswana and the intensity of HIV/AIDS infection is represented using a colour scale where lower values are represented in green and the higher values in light purple. The natural neighbour map shows one district that has highest HIV/AIDS prevalence rate. The prevalence rate decreases as we move to the western part of the country. In IDW map shows one district (Selebi-Phikwe) with a prevalence rate higher than 16.3%.

The prevalence rate decreases towards the western part of the country. In ordinary kriging map HIV/AIDS is higher in Selebi-Phikwe, Central-Bobonong and Central-Mahalapye. Like other methods, the prevalence decreases as we move to the western part of the country. Generally the maps show a wide geographical disparity of HIV/AIDS with the north-eastern and central districts having higher prevalence compared to the western districts.

4.4.3. Comparison of interpolation methods

In this section the effectiveness of inverse distance weighting, natural neighbour and ordinary kriging interpolation methods are examined. The root mean square error (RMSE used to determine the effectiveness of interpolation methods. RMSE for IDW, NN, OK was calculated in ArcGIS 10.1 (Johnston, *et al.*, 2001) and the results s are shown in Table 4.11

Table 4.11 Showing RMSE for the three interpolation methods

Interpolation method	RMSE	Variance
Ordinary Kriging	6.3263	-0.2085
Nearest neighbour	6.84154	0.321723
Inverse distance weighting	7.1812	0.5394

Table 4.11 results indicate that of the interpolation methods used to determine the spatial distribution of HIV/AIDS in Botswana, ordinary kriging was the best method. Generally, the RMSE value of the stochastics technique (Ordinary Kriging) was lower than for the deterministic technique (Inverse distance weighting and Natural Neighbour). Ordinary kriging interpolation proved to be most likely to produce the best spatial continuous surface for HIV/AIDS prevalence.

4.5. Logistic Regression results

In this section logistic regression was employed to determine risk factors that are associated with high prevalence rate of HIV/AIDS in some districts of Botswana. The variables that were considered are gender, education, marital status, employment, and place of residence, religious affiliation, circumcision status, and use of condoms, use of alcohol, and number of sexual partners, knowledge and age at first marriage. Odd ratio (OR), a measure of association between an independent variable and dependent variable was used to interpret regression.

Odds ratio

The odds ratio (OR) is a comparative measure of two odds relative to different events. For two events A and B, according to Park, 2013 the corresponding odds of A occurring relative to B occurring is

$$\text{Odds ratio } |\text{A vs B}| = \frac{\text{odds}|\text{A}|}{\text{odds}|\text{B}|} = \frac{P_A/1-P_A}{P_B/1-P_B} \qquad (22)$$

OR represents the odds that an outcome (e.g. HIV/AIDS) will occur given a particular exposure (e.g. alcohol abuse, condom use) compared to the odds of the outcome occurring in the absence of that exposure. It can be used to establish if a particular exposure is a risk factor for a particular outcome and compare the significance of various risk factors for that outcome (Parks, 2013). If OR = 1, it indicates that the exposure does not affect odds of outcome, OR >1 indicates exposure associated with high odds of outcome and OR < 1indicates exposure associated with lower odds of outcome. OR is the exponential function of the regression coefficient (e^{b_1}).

This tells us that β_1 represents the change in the odds of an outcome for an increase in one unit of X (the independent variable). A positive value of β_1 indicates that the odds would increase while a negative indicates that the odds would decrease.

Logistic regression was performed for each of the selected districts of this study and the significant factors were determined using the p-value and Wald test. The likelihood ratio test was used to determine the goodness of fit for the models. Only the significant risk factors were reported and the results are summarised in Tables 4.12 through 4.17.

Risk factors associated with HIV/AIDS in Gaborone district.

The logistics was statistical significant with a $\chi^2_{(7)}$= 3917.667 and a p-value less than 0.005.The model explained 54.7 % ((R^2) of the variance of HIV/AIDS and correctly classified 97.4% of the cases.

Table 4.12 Risk factors associated with HIV/AIDS in Gaborone district

Variable		B	Wald	Sig.	Exp(B)	95% CI for EXP(B)	
						Lower	Upper
Education	Higher (Ref)		304.126	.000			
	Primary	.536	6.893	.009	1.709	1.145	2.549
	Junior Sec	7.758	298.265	.000	2339.81	970.114	5643.406
	Senior Sec	.838	14.689	.000	2.311	1.506	3.548
Circumcision	No(Ref)	-	157.624	.000	-	-	-
	Yes	-4.385	256.937	.000	.012	.007	.021
Age	15-19	-	459.881	.000	-	-	-
	25-29(Ref)	-2.084	34.622	.000	.124	.062	.249
	30-34	1.690	30.637	.000	5.420	2.979	9.862
	35-39	3.043	104.307	.000	20.972	11.695	37.607
	40-44	5.140	234.713	.000	170.662	88.425	329.381
Condom use	Yes (Ref)	-	550.947	.000	-	-	-
	Never	5.811	449.624	.000	333.837	195.109	571.202
	Sometimes	-.788	25.046	.000	.455	.334	.619
No of sex partners	1(Ref)	-	30.51	.000	-	-	-
	2 or more	2.235	50.199	.000	9.347	5.037	17.346
Alcohol	No(Ref)	-	286.245	.000	-	-	-
	Yes	4.060	239.669	.000	57.985	34.680	96.951

Table 4.12 shows the variables, condom use, number of sexual partners, age at first marriage, circumcision and education were significant risk factors that increased the prevalence rate of HIV/AIDS in Gaborone. Having primary education was not protective against HIV/AIDS infection as it increased by 70.9% [OR = 1.709, 95% CI (1.145-2.549), p -value= 0.000] and the odds of those with junior secondary education was significantly high [OR = 2339.82, 95% CI (970.11-643.4), p= 0.000].

Older individuals have a higher chance of being infected by HIV/AIDS as compared younger individuals. For instance, the odds of being HIV positive increased 5.4 times [OR = 5.42, 95% CI (2.979-9.862), p= 0.000], 20.97 times [OR=20.97, 95% CI 911.695-37.607), p=0.000] and 170.66 times [OR=170.66 95% CI (88.425 – 329.381), p=0.000] when the participant's age was between the ages of 30-34 years 35-39 years and 40-44 years respectively. Moreover, not using condoms during sexual intercourse increased the chance of being HIV positive by 333.84 [OR=333.84, 95% CI (195.11-571.20), p=0.000]. Using condoms sometimes had increased the chance of getting HIV/AIDS infection by 0.455 [OR=0.455, 95% CI (0.334-0.619), p=0.000]. The chance of being HIV positive increased for individuals who engaged in multiple sexual relationships [OR=9.347, 95%CI (5.04-17.35), p=0.000], and drinking alcohol were 58 times [OR= 57.99, 95% CI (34.68-96.95), p=0.000] more likely to get HIV/AIDS infection than those who did not drink

Risk factors associated with HIV/AIDS Prevalence in Francistown

Table 4.13 gives a summary of logit model for risk factors of HIV/AIDS in Francistown.

Table 4.13 HIV/AIDS factors in Francistown

Variable		B	S.E.	Wald	Sig.	Exp(B)	95% C.I.for Exp(B)	
							Lower	Upper
Education	None(Ref)	-	-	208.743	.000	-	-	-
	Primary	.720	.131	30.372	.000	2.055	1.590	2.654
	Junior secondary	4.667	.354	173.725	.000	106.406	53.156	213.000
	Senior secondary	.856	.147	33.860	.000	2.353	1.764	3.139
Circumcision	No(Ref)	-	-	57.324	.000	-	-	-
	Yes	-.582	.145	16.140	.000	.559	.421	.742
Age	15-19 (Ref)	-	-	251.698	.000	-	-	-
	30-34	-2.460	.241	104.483	.000	.085	.053	.137
	35-39	.584	.235	6.172	.013	1.793	1.131	2.842
	40-44	-1.035	.198	27.452	.000	.355	.241	.523
Sexual partners	1(Ref)	-	-	214.364	.000	-	-	-
	>2	1.692	.127	176.918	.000	5.431	4.232	6.969
Alcohol	No(Ref)	-	-	162.895	.000	-	-	-
	Yes	1.937	.133	210.727	.000	6.940	5.343	9.014

In Francistown, the variables age, education, number of sexual partners, circumcision and alcohol had significant effect on the prevalence of HIV/AIDS. Individuals who had primary education were 2.1(OR= 2.055, 95% CI 1.59-2.65(), p=0.001], those with junior secondary education were 106.4 [OR= 106.4 95% CI= (53.16-213.00), p=0.00] and those with senior secondary were2.4 (OR=2.35, 95% CI (1.76-3.14), p=0.00 times more likely to be HIV positive as compared to those who had no education. Those who took alcohol were 6.9 times likely to be HIV positive as compared to those who did not take alcohol. Individuals who had a multiple partners were 5.3 times more likely to be HIV positive. Individuals in ages 30-34 years, 35-39 years and 40-44 years were 0.085, 1.79 and 0.36 times more likely to be HIV positive than those in the ages 15-19 years.

Risk factors associated with HIV/AIDS prevalence in Selebi-Phikwe, Central Bobonong, Central Boteti, Kagtleng, Southeast and Ngamiland South

The regression logistic model explained all the variations of HIV/AIDS with R^2 = 100% and correctly classified 92.9% of the cases. Selebi-Phikwe, Central Bobonong, Central Boteti, Kagtleng, Southeast and Ngamiland South had similar risk variables. These were employment, age at first marriage, number of sexual partners, use of condoms and alcohol. The results are presented in Table 4.14.

Table 4.14 Risk factors associated with HIV/AIDS prevalence in Selebi-Phikwe, Central Bobonong, Central Boteti, Kagtleng, Southeast and Ngamiland South

Variable		B	S. E.	Wald	Sig.	Exp(B)	95% C. I. Exp(B)	
							Lower	Upper
Employment	Full-time(Ref)	-	-	7.043	.217	-	-	-
	Self-employed	.505	.190	7.040	.008	1.658	1.141	2.408
Age at first marriage	12-19(Ref)	-	-	129.116	.000	-	-	-
	20-29	-4.346	.384	127.829	.000	.013	.006	.028
	30-39	-5.339	.489	119.141	.000	.005	.002	.013
Alcohol use	No(Ref)	-	-	286.245	.000	-	-	-
	Yes	-4.060	.262	239.669	.000	.017	.010	.029
Condom use	No(Ref)	-	-	550.947	.000	-	-	-
	Yes	-6.599	.283	541.887	.000	.001	.001	.002
	Sometimes	-5.811	.274	449.624	.000	.003	.002	.005
Sex Partner	>2	-	-	34.584	.000	-	-	-
	1	-2.235	.315	50.199	.000	.0.107	.058	.199

According to the findings, being self-employed increased the risk of being HIV positive by 65.8%; [OR = 1.6508, 95% CI (1.14-2.41), p= 0.000] as compared to being a full-time employee. Delaying marriage in Selebi-Phikwe, Central Bobonong, Central Boteti, Kgatleng, Southeast and

Ngamiland South districts reduced the chance of being HIV positive by 0.013 [OR = 0.013, 95% CI (0.006-0.028), p=0.000] and 0.005 [OR = 0.005, 95% CI (0.002-013), p=0.00] times in the ages 20-29 years and 30-39 years compared to those who married when they were young.

Having a single sexual partner reduced the chance of being HIV positive by 0.107 [OR = 0.107, 95% CI (0.058-0.199), p = 0.000]. The risk of HIV/AIDS infection was reduced by 98.3% [OR=0.017, 95% CI (.010 - .029) p=0.000] in respondents who did not drink alcohol. Using condoms is very critical in the prevention of HIV/AIDS and other risk factors such as sexually transmitted infections. The chance of getting infected by HIV/AIDS was decreased by 0.001[OR =0.001, 95% CI (0.001-0.002), p = 0.000] for those who used protection always and 0.003 [OR = 0.003, 95% (0.002 -0.005), p= 0.000] for those who sometimes used condoms compared to those who never used condoms.

Risk factors associated with HIV/AIDS prevalence in Kweneng East District

A logistic regression was performed to determine the factors that contribute to the spatial distribution of HIV/AIDS in Kweneng East district. The logistic regression was statistical significant with a $\chi^2_{(7)}$= 1033.196 with a p-value less than 0.005. The model explained 65.6% (R^2) of the variance in HIV/AIDS and correctly classified 99.3% of the cases.

Kweneng East district is an urban village and is composed of Molepolole, Thamaga, Gabane, Mogoditshane and Kopong villages. Education, age at first marriage, number of sexual partners, use of condoms and alcohol were significantly associated with the prevalence of HIV/AIDS in Kweneng East district. The results are presented in Table 4.15.

Table 4.15 Risk factors associated with HIV/AIDS prevalence in Kweneng District

Variable		B	S. E.	Wald	Sig.	Exp(B)	95% C. I. Exp(B)	
							Lower	Upper
Education	Higher(ref)	-	-	304.126	.000	-	-	-
	Junior Sec	7.758	.449	298.265	.000	2339.81	970.114	5643.406
	Senior Sec	.838	.219	14.689	.000	2.311	1.506	3.548
Age at first marriage	12-19(Ref)	-	-	129.116	.000	-	-	-
	20-29	-4.346	.384	127.829	.000	.013	.006	.028
	30-39	-5.339	.489	119.141	.000	.005	.002	.013
Alcohol use	No(Ref)	-	-	286.245	.000	-	-	-
	Yes	4.060	.262	239.669	.000	57.985	34.680	96.951
Condom use	No(Ref)	-	-	550.947	.000	-	-	-
	Yes	-6.599	.283	541.887	.000	.001	.001	.002
	Sometimes	-5.811	.274	449.624	.000	.003	.002	.005
No of Sex Partners	>2(Ref)	-	-	34.584	.000	-	-	-
	1	2.235	.315	50.199	.000	9.347	5.037	17.346

Individuals who had junior and senior secondary education were 2339.82 [OR=2339.8, 95% CI (970.11-5643.4), p=0.000] and 2.31 [OR=2.311, 95% CI (1.506-3.548), p=0.000] times more likely to be HIV positive compared to those with higher education. Getting married when an individual is older reduced the chance of being HIV positive by 0.013 and 0.005 in the ages 20-29 years and 30-39 years respectively. Individuals with a single partner were 9.35 times more likely to be HIV positive compared to those with multiple partners. Using condoms always reduced the chance of acquiring HIV/AIDS infection by 0.001 [OR=0.001, 95% CI (90.001 0-.002), p=0.000] and 0.003 [OR=0.003, 95%CI (0.002-0.005), p=0.000] for those who sometimes use condoms.

Risk factors associated with HIV/AIDS in Central Serowe and Central Tutume

The logistic regression model was statistical significant with $\chi^2_{(7)}$= 3917.667 and p< 0.005. 65.6 % (R^2) variance in HIV/AIDS has been explained by the model and 99.2% of the cases were correctly classified. In Central Serowe and Central Tutume, employment, age at first marriage, alcohol and use of condoms were risk factors associated with HIV/AIDS infection. Summarised in Table 4.16 are the results of logit model for risk factors of these districts.

Table 4.16 Risk factors associated with HIV/AIDS in Central Serowe and Central Tutume

Variable		B	S. E	Wald	Sig.	Exp(B)	95% C. I. Exp(B)	
							Lower	Upper
Employment	Fulltime(ref)	-	-	37.882	.000	-	-	-
	Self-employed	1.033	.168	37.880	.000	2.808	2.021	3.902
Age at first marriage	12-19(Ref)	-	-	159.546	.000	-	-	-
	20-29	-4.420	.350	159.282	.000	.012	.006	.024
	30-39	-5.084	.445	130.757	.000	.006	.003	.015
Alcohol use	No	-3.084	.196	246.850	.000	.046	.031	.067
Condom use	No(Ref)	-	-	664.927	.000	-	-	-
	Yes	-6.335	.248	651.430	.000	.002	.001	.003
	Sometimes	-5.484	.242	512.049	.000	.004	.003	.003

In Central Serowe and Central Tutume, individuals who were self-employed were 2.8 times more likely to be HIV positive compared to those who were employed on full-time basis. Delaying marriage reduced the chance of acquiring HIV/AIDS by 0.012 [OR=0.012, 95%CI (0.006-0.024), p=0.000] for individuals in the ages 20-29 years and by 0.006 [OR=0.006, 95%CI (0.003-0.015), p= 0.000] for individuals in the ages 30-39 years. Not taking alcohol reduced the effect of acquiring HIV/AIDS infection by 0.046 [OR=0.046, 95%CI (0.031–0.067), p=0.000]. Respondents who always used condoms and those who sometimes used them had their chance of acquiring of HIV/AIDS infection reduced by 0.002 [OR=0.002, 95%CI (0.001 – 0.003), p=0.000] and 0.0004 {OR=0.0004, 95%CI (0.003–0.007), p=0.000].

Risk factors associated with HIV/AIDS Central Mahalapye district

The logistic regression was statistical significant with $\chi^2_{(7)} = 3917.667$ and $p<0.005$. The model explained 65.6% (R^2) and classified 99.3% of the cases accurately. Education, age, age at first marriage, number of sex partners, alcohol and use of condoms were risk variables that were associated with HIV/AIDS infection in Central Mahalapye and the results are summarised in Table 4.17.

Table 4.17 Risk factors associated with HIV/AIDS Central Mahalapye district

Variable		B	S. E.	Wald	Sig.	Exp(B)	95% C. I. Exp(B)	
							Lower	Upper
Education	Higher(ref)	-	-	304.126	.000	-	-	-
	Primary	.536	.204	6.893	.000	1.71	1.15	2.55
	Junior Sec	7.758	.449	298.265	.000	2339.817	970.114	5643.41
	Senior Sec	.838	.219	14.689	.000	2.311	1.506	3.55
Age	15-19(Ref)	-	-	542.718	.000	-	-	-
	25-29	-2.08		459.881	.000	.12	0.062	0.249
	30-34	1.69	.354	34.622	.000	5.42	2.98	9.86
	35-39	3.04	.305	30.637	.000	20.97	11.69	37.61
	40-44	5.14	.305	30.637	.000	170.66	88.42	329.38
Age at first marriage	12-19(Ref)	-	-	129.116	.000	-	-	-
	20-29	-4.346	.384	127.829	.000	.013	.006	.028
	30-39	-5.339	.489	119.141	.000	.005	.003	.013
Alcohol use	Yes	4.060	.262	239.669	.000	57.985	34.86	96.9
Condom use	Yes(Ref)	-	-	550.947	.000	-	-	-
	No	5.811	.274	449.624	.000	333.837	196.11	577.24
	Sometimes	-.788	.157	25.046	.000	.455	.334	.619
Sex Partner	>2(Ref)	-	-	34.584	.000	-	-	-
	1	2.235	.315	50.199	.000	9.347	5.037	17.35
Place	Cities(Ref)	-	-	171.514	.000	-	-	-
	Urban village	2.040	.237	74.073	.000	7.689	4.832	12.236
	Rural	2.898	.226	165.043	.000	18.143	11.659	28.231
Circumcision	No(Ref)	-	-	157.249	.000	-	-	-
	yes	-4.39	.274	256.937	.000	.012	.007	.021

The chance of acquiring HIV/AIDS for individuals with primary education was increased by 71% [OR=1.17, 95%CI (1.15-2.55) p=0.000]. Individuals with junior and senior secondary education were 2339.82 and 2.31 times more likely to be HIV positive compared to those with higher education. Table 4.17 shows that the risk of being HIV positive for respondents aged 25-29 years was decreased by 0.12 [OR = 0.12, 95% CI (0.062-0.249), p=0.000]. Participants in the ages 30-34 years, 35-39 years and 40-44 years were 5.42, 20.97 and 170.97 times more likely to be HIV positive compared to those in the ages 15-19 years. They were benefits in delaying marriage. For instance, getting married when your age was between 20-29 years reduced the chance of having HIV/AIDS infection by 0.013 [OR=0.013, 95%CI (0.006-0.028), p=0.000] and by 0.005 [OR=0.005, 95%CI (0.003-0.013), p=0.000] when your age was between 30-39 years.

Not using of condoms during sexual intercourse increased the chance of acquiring HIV/AIDS infection by 333.84 while using condom sometimes decreased the chance of HIV/AIDS infection by 0.46 [OR=0.46, 95%CI (0.33-0.61), p=0.00]. Being circumcised had a benefit as it decreased the chance of having HIV/AIDS infection by 0.01. Individuals who had a single sexual partner were 9.35 times more likely to be HIV positive compared to those with multiple sexual partners. Participants living in urban villages and in rural villages had their chance of being HIV positive increased by 7.69 [OR=7.69, 95%CI (4.83-12.23), p=0.000] and 18.14 [OR=18.14, 95%CI (11.66-28.23), p=0.000] respectively.

Assessment of Goodness-of-fit

One of the objectives of the study is to identify risk factors that contribute to the spatial distribution of HIV/AIDS in selected districts of Botswana. Logistic regression helped to identify risk factors of HIV/AIDS for each of the selected districts and a Wald test was used to test the significance of each individual risk factor. Goodness- of-fit was assessed and the results are summarised in Table 4.18.

Table 4.18 -2Loglikelihood ratio test and R^2

District	-2Loglikelihood	R^2
Gaborone	3954.848	0.547
Francistown	3039.789	0.656
Kweneng East	3039.789	0.656
Central Mahalapye	3039.789	0.656
Central Serowe and Tutume	3039.789	0.656
Selebi-Phikwe and others	3039.789	0.656

The variables included in the study were significant (p=0.000). Also calculated was the likelihood test to compare the reduced model from the full model. The result of the likelihood ratio test for Gaborone was 3954.848 with R^2= 0.547 and 3039.789, with R^2= 0.656 for all the other districts. The regression models for Tables 4.12 through to 4.17 are presented as follows:

HIV Prevalence(Gaborone) = 0.361 + 0.536Edu(primary) + 7.758Edu (Jnr) + 0.838Edu (Snr) + 2.04Place(town) + 2.898Place (urban) – 2.084Age(25-29) + 1.69Age(30-34) + 3.043Age(35-39)+5.14Age(40-44) + 5.811Condom use(Never) - 0.788Condom use(sometimes) + 2.235Numsx(2 or more) + Alcohol use(Yes) (23)

HIV Prevalence (Francistown) = 0.361 – 1.096 Age (30-34) + 0.694Age (35-39) – 0.428Age (40-44) + Agma (20-29) + Agma (30-39) + 1.209Alcohol use + 1.282Numsx (1) (24)

HIV Prevalence (Selebi-Phikwe and others) = 0.361 + 0.505Employment (self) – Agma (20-29) – 5.339$Agma_2$ (30-39) - 4.06Alcohol use (Yes) – 6.599Condom use (Yes) -5.811Condom use (sometimes) -2.235 Numsx (1) (25)

HIV Prevalence (Kweneg East) =0.361 + 7.758Edu (Jnr) + 0.838Edu (Snr) – 4.346Agma (20-29) -5.339Agma (30-39) + 4.06Alcohol use (Yes) + 2.35numsx (1) (26)

HIV Prevalence (Central-Tutume and Central-Serowe) = 0.361 + 1.033Employment (Self) – 4.42Agma (20-29) – 5.084Agma (30-39) -3.084Alcohol use (No) –Condom use (Yes) – 5.484Condom use (sometimes) (28)

HIV Prevalence(Central-Mahalapye) = 0.361 + 0.536Edu (primary) + 7.758Edu(Jnr) + 0.838Edu (Snr) – 2.08Age(25-29) + 1.69Age(30-34) + 3.04Age(35-39) + 5.14Age(40-44) – 4.346Agma(20-29) – 5.339Agma(30-39) + 4.06Alcohol use(Yes) + 5.811Condom use(No) – 7.88Condom use(sometimes) + 2.235Numsx(1) + 2.04Place(urban) +2.898Place (rural) – 4.39Circumcison (Yes) (29)

where

Agma=Age at first marriage, Nusx=number of sexual partners, Place= place of residence, Edu=education, Jnr= Junior, Snr= senior

The model for Gaborone (equation 1) accounted for 54.7% of the variance, while the models for all the other districts accounted for 65.6% of the variance. This implies that the reduced model is a good model for predicting the prevalence of HIV/AIDS.

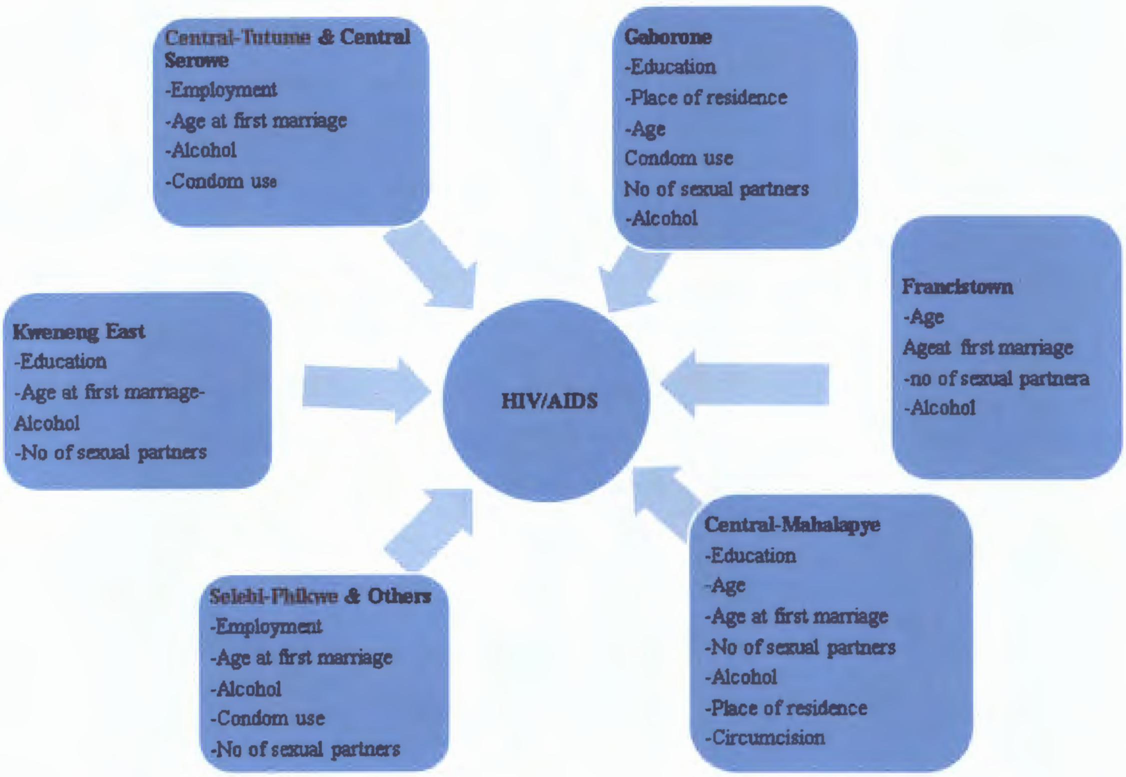
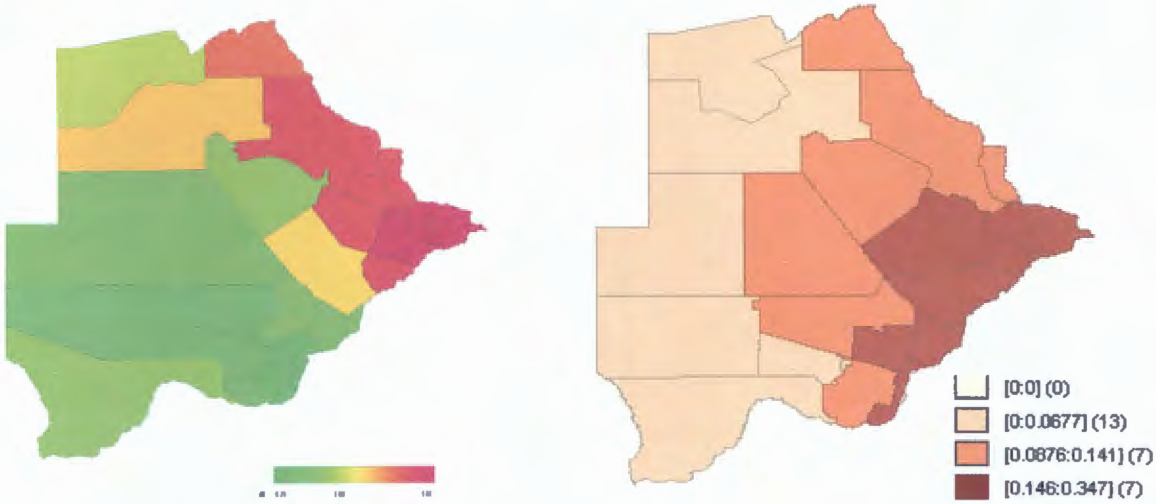


Figure 4.8 factors contributing to the spatial distribution of HIV/AIDS

4.6. Comparison with previous studies

The study findings are similar to those obtained by Kandala *et al.* (2012). In their study, the authors looked at the geography of HIV/AIDS in Botswana using data from 2008 BISA III. Bayesian geo-additive mixed models based on Markov Chain Monte Carlo techniques were used to map the geographic distribution of HIV in 26 districts of Botswana. They produced age and location adjusted maps. The current study employed GIS techniques and 2013 BIAS IV data to study the spatial distribution of HIV/AIDS in 12 highly populated districts of Botswana. In particular it looked at spatial autocorrelation; clusters of HIV/AIDS, and it determined the best interpolation method and identified risk factors associated with spatial distribution.

Figure 4.14 shows the spatial distribution of HIV/AIDS for the two studies.



Spatial map from previous study Spatial map from current study

Figure 4.9shows spatial distribution of HIV/AIDS from the previous study and the current study.

Both studies indicate significant and clear spatial distribution of HIV/AIDS. In the study by Kandala *et al.*, 2012 HIV/AIDS is higher in the north-eastern districts (Francistown Selebi-Phikwe and North-east) and the intensity reduces towards the northern districts. There is a sharp diving line running through the central districts. HIV/AIDS is higher in the northern parts of these districts and lower in the western parts. For the current study HIV/AIDS is higher in the southern, northern and the central districts. The intensity reduces towards the western parts of the country.

It can be observed from the two maps that there is a shift in the concentration of HIV/AIDS. The current study from the current data indicated a south, north-east and central concentration of HIV/AIDS. The findings from this study will help government and service providers to reallocate HIV/AIDS prevention resource to the districts where they are needed most and still support basic prevention for everyone across the country.

4.7. Summary

In this chapter the distribution of HIV/AIDS, spatial analysis, interpolation and logistic regression were carried out. The result indicated that HIV/AIDS was higher in Selebi-Phikwe, Francistown and Central Mahalapye districts and lower in Ngamiland South and Central Serowe. Autocorrelation results showed that HIV/AIDS was randomly distributed and one primary cluster was detected. Interpolation was carried out using three methods, Inverse distance weighting (IDW), Natural neighbour (NN) and Ordinary kriging (OK) and Ordinary kriging was reported to be the best interpolation technique.

Logistic regression analysis was employed to determine risk factors associated with HIV/AIDS. Some district had few risk factors associated with HIV infection, for example Selebi-Phikwe, Central Bobonong, Central Boteti, Kgatleng, Southeast and Ngamiland South district had only four risk factors (number of sexual partners alcohol use, age at first marriage and use of condoms) while Mahalapye and Gaborone. Age at first marriage, alcohol use, number of sexual partners and condom use were the most common risk factors for all the selected districts of Botswana.

Both previous studies and the current study show that there is a significant spatial distribution of HIV/AIDS in Botswana and that it is higher in the north eastern district. The current study shows south, central and north east spatial distribution while the previous shows a north- north east distribution

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1. Introduction

This chapter reflects on the findings of the current study and its effects on further study. Conclusions drawn in this study are in relation to the objectives listed in chapter one. The findings from this study are compared to those of previous studies noting similarities and differences. Recommendations for further research on spatial distribution of HIV/AIDS using different interpolation methods and also inclusion of all districts in Botswana are provided in this chapter.

The aim of the study was to investigate spatial distribution of HIV/AIDS using different methods and comparing their performance. The study was motivated by the fact that there is limited study on spatial distribution of HIV/AIDS using GIS and no study has compared the performance of different spatial interpolation methods in Botswana. Also no study has unveiled why some districts have high/ low HIV/AIDS prevalence than others.

The investigation was done using 2013 Botswana Aids Impact Survey IV (2013 BAIS IV) data collected by Central Statistics Office Botswana in collaboration with National Aids Coordinating Agency (NACA) and the Ministry of Health. The study population was HIV positive individuals between the ages 15-19 years. 12 districts with a population above five thousands were selected for the study. Quantitative descriptive statistics were used to view the general spread of HIV/AIDS. Autocorrelation analysis, in particular Moran's I and Getis-Ord was employed to determine the similarities and dissimilarities of HIV/AIDS characteristics in neighbouring districts. Hotspot analysis was done using LISA, Getis-Ord Gi* and Kulldorff's scan statistics to identify clusters with high-high, low-low, and high-low and low-high HIV/AIDS prevalence. Continuous surface maps were created using IDW, NN and OK spatial interpolation methods and their performances compared. Logistic regression analysis was conducted to determine risk factors that are contributed to the spatial distribution of HIV/AIDS in Botswana and spatial maps from the current study were compared with those from previous studies. The findings of this study have provided current spatial distribution of HIV/AIDS and reasons why some districts have high/low HIV/AIDS prevalence in Botswana.

5.2. Conclusions

In this section important findings are highlighted with reference to the research questions. It aims to confirm whether the objectives of this study have been achieved. The section also highlights the advances in knowledge in this area of study.

RESEARCH QUESTION 1

Do districts sharing common boundaries have similar or dissimilar HIV/AIDS characteristics?

Conclusion 1

In order to answer this question, the study performed autocorrelation analysis. In particular Moran's I and Getis-Ord General G was used to measure the extent of correlation between neighbouring districts. Numerous researchers, for example Moise and Kalipeni (2010), Hsueh *et al.* (2012), Obidoa and Cromley (2012) and Zulu *et al.* (2014) have used the same method.

The findings from both statistics indicate positive correlation between neighbouring districts. The results suggest that proximity of districts may have an association to the diffusion of HIV/AIDS between neighbouring districts. These findings are in agreement with findings from previous studies that showed that nearer objects have similar characteristics (Kwandal *et al.*; 2009, Peng *et al.*; 2011, Obidoa and Cromley, 2012; Moise and Kalipeni, 2010 Zulu *et al.*; 2014, Kandala *et al.*; 2012, Zhou *et al.*, 2014; Zhang *et al.*, 2015). Moran's I and Getis-Ord general G statistics successfully examined the extent of clustering that existed in the selected districts. A positive Moran's I value indicated that spatial clustering of HIV/AIDS is more prominent in Botswana. On the other hand a high z-score in Getis-Ord also indicates spatial clustering.

RESEARCH QUESTION 2;

Which districts are hotspots or cold spots?

Conclusion 2

To answer this question, the study employed Local indicator of spatial analysis (LISA), Kulldorff SaTScan, and Getis-Ord Gi* statistics to pinpoint clusters in the study area. The results are presented in Figure 4.9 and Figure 4.10 for LISA and Getis-Ord Gi*and Table 4.10 and Table 4.11 for Kulldorff Sat Scan.

The conclusions made are based on the results in chapter 4 created using Geo-Da. The maps showed one district that had high-high HIV/AIDS prevalence and this was Central-Serowe. There were two districts that had low-low HIV/AIDS prevalence and these were Kgalagadi south and Ngwakwetse. These results are significant at 5% significance level. Similar results are depicted in Figure 4.10(a) with one more district in the low-low category. The results are significant at 5% and 1% for high-high and low-low categories respectively. One primary cluster and three secondary clusters were identified (Tables 4.10 and 4.11). Central Serowe was the only primary cluster identified by Kulldorff scan with a relative risk of 7.6. High-high in this study indicated that the district has high HIV/AIDS prevalence and is surrounded by districts which also have high HIV/AIDS prevalence. Similarly low-low indicated districts that have low HIV/AIDS prevalence and are surrounded districts with low HIV/AIDS prevalence.). Of the three secondary clusters, Francistown and Selebi-Phikwe were significant secondary clusters with a risk of 1.16 and 1.36 respectively. In Central-Mahalapye, the risk of getting HIV/AIDS inside the district was the same as outside the district (risk = 1.01) hence it not a significant cluster.

From the results, it can be concluded that Central-Serowe is the epicentre of HIV/AIDS in Botswana. The district is the largest village in Botswana and one of the largest in Africa (Botswana Beckons). Central-Serowe is located along the main road to Orapa mine and near Botswana Defence Force (BDF) camp. Mining is largely a male occupation (Kandala *et al.,* 2012). Like all other mines, Orapa mine is likely to employ males mostly and these are likely to be from Central-Serowe and other neighbouring districts. These men are often separated from their families and are likely to engage in risky sexual behaviour. This could account for the high HIV/AIDS prevalence in the district.

Previous studies have associated high HIV/AIDS infection to proximity to the main road (Messina *et al.*, 2010; Kalipeni and Zulu, 2008; Zulu et al., 2014). Getis-Ord Gi* revealed two districts (Central-Mahalapye and Central-Serowe) that had high concentration of HIV/AIDS and three districts (Selebi-Phikwe, Central Bobonong and Francistown) that were of moderate clustering. There were three districts that were cold spots in the study area and these are Ngwakwetse, Kgalagadi north and south. The findings from this study show a variation in the methods used to pinpoint the clusters. LISA statistics and Kulldorff scan showed one district (Central-Serowe) as the hotspot while Getis-Ord Gi* showed two districts (Central-Serowe and Central-Mahaplapye).

RESEARCH QUESTION 3

Which is the best interpolation method for BIAS IV?

Conclusion 3

In this section three interpolation methods were used and compared to get the best interpolation technique for BAIS IV data. The results are shown in Figure 4.12 and Table 4.13.

Prior to the analysis, exploratory spatial analysis data analysis was carried out to check if data was normally distributed and that it had no trend. The histogram (Figure 4.7) showed that data was normal distributed with mode and median almost equal (12.5 and 12.4) respectively. Trend analysis was carried to test for trend and the results in Figure 4.8 revealed that data had a strong trend. For details, refer to section 4.3.2 in chapter 4. The trend indicates that HIV/AIDS is concentrated in the central districts and decreases towards the western parts of the country. Data was transformed using second order polynomial so that kriging interpolation technique could be used.

There are many kriging methods; hence there was a need to choose the best. Three kriging methods were employed and a cross validation procedure was carried out to determine the best method (see Section 3.5.5). Ordinary kriging was chosen as it had the least errors (Table 4.12) and is the most popular method (Webster *et al.*, 1994; Zandi *et al.*, 2011; Zulu *et al.*, 2014).

In this study the performance of three interpolation methods was evaluated. Raster surfaces (Figure 4.12) for the three interpolation methods were extracted to estimate values of HIV/AIDS data for the selected districts. It should be noted that no interpolation methods stand out to be superior by a large margin and that several interpolation methods perform best when specific criterion is considered (Hofierka *et al.*, 2007). Inverse distance weighting, Ordinary kriging and Natural neighbour interpolation methods were used and their performance was measured using the RMSE and variance.

The results for variance and the RMSE are shown in Table 4.13. The findings revealed that the IDW technique performed worse in terms of its variability 0.5394 and 7.1812 for the variance and RMSE. Ordinary kriging (OK) showed high accuracy considering the least variance (0.2085) and RMSE (6.326). Natural Neighbour (NN) fell in between with variance 0.321 and RMSE 6.841. The results obtained from the comparison of the three interpolation methods indicated that OK was the most suitable method for mapping the spatial distribution of HIV/AIDS for this study. BAIS IV is a nationally representative demographic survey with large samples. The finding supports the assumption that ordinary kriging performs better for large samples (Johnston *et al.*, 2001). This study contributes to the body of knowledge in that it is one of first studies to compare performance of interpolation methods using current HIV/AIDS data. The study also contributes to the body of knowledge by using GIS to analyse spatial distributions which is limited in Botswana.

RESEARCH QUESTION 4

Which factors contribute to the spatial distribution of HIV/AIDS?

Conclusion 4

This subsection used logistic regression to identify risk factors associated with the spatial distribution of HIV/AIDS in the selected districts of Botswana and this answered the question why some districts have high or low HIV/AIDS prevalence than others. Logistics regression has many advantages, for instance it is more robust and independent variables do not have to be normally distributed, or have equal variance in each group. The results from this method are shown in Table 4.14 through to Table 4.19.

This study has successfully identified risk factors that were associated with HIV/AIDS in the selected districts. These were determined by the Wald test and p-value (<0.05). Variables that were determined as significant risk factors were education level, circumcision, and age, and condom use, number of sex partners, alcohol use and age at first marriage. Of the significant variables, condom use, number of sexual partners and alcohol use were common to all the districts.

In Botswana, alcohol is seen as a form of entertainment (Seloilwe, 2005). Excessive intake of alcohol reduces judgment and leads to promiscuous, irresponsible and high risk sexual behaviour. Also, individuals who drink a lot tend to engage in casual and unprotected sex which increases the risk of contracting HIV/AIDS. The study found age at first marriage to be a significant risk factor in all the selected districts. Individuals who delayed getting married were at risk of HIV/AIDS infection. One of the reasons why individuals who delayed marriage were at risk of getting infected with HIV/AIDS could be that they were likely to have engaged in sexual activities with several partners before marriage and were more likely to be infected

Condom use is one of the risk factors and it remains low across the country (Avert, 2014). Percentage of people using condom has decreased from 90.2% in 2008 to 81.9% in 2012 (Avert, 2014). One of the reasons why condom use remains a challenge in Botswana is that the country is struggling to dispel the myth and views surrounding the prevention of HIV/AIDS and transmission with cultural beliefs. The findings are consistent with other studies. Eaton *et al.* (2003) in their study indicated that low use of condom may be attributed to the fact that it was reserved for prostitutes, untrustworthy partners and casual partners. The belief that sex is more pleasurable without condoms and the general understanding of HIV/AIDS maybe one of the reasons why there was low use or no use of condom

Circumcision is associated with HIV/AIDS infection in Gaborone and Francistown. This is contrary to other studies carried out in South Africa, Kenya and Uganda that concluded that male circumcision significantly reduces the chance that men will contract HIV/AIDS through intercourse with infected women by up to 70% (Auvert *et al.*, 2005, William et al., 2006, Bailey et al., 2007, Gray *et al.*, 2007, Mavedzenge *et al.*, 2011). Education level, circumcision and age were significant risk factors in Gaborone, Francistown and Central Mahalapye. For the variable education level, junior secondary was the most critical risk factor with odds ratio 2339.8 in Gaborone and Central Mahalapye and odds ratio 106.4 in Francistown.

Age was a risk factor associated with HIV/AIDS in Gaborone, Francistown and Central Mahalapye. Individuals in the age groups 25-29 years, 30-34 years, 35-39 years and 40-44 years were 5.42 times, 20.97 times and 170.66 times more likely to be HIV positive as compared to those in the age group 15-19 years. The risk of getting infected was increasing with age in Gaborone and Central Mahalapye. These findings are in contrast to findings from other studies such as those conducted in Tanzania and Uganda. It was found that older age is a protective factor against HIV/AIDS infection (Magadi and Desta, 2011).

This study found that the number of sex partners to be significantly associated with HIV/AIDS in most selected districts of Botswana except Central-Serowe and Central-Tutume. These findings are consistent with findings from other studies (Uchudi *et al.*, 2012). To my knowledge this is the first study that has attempted to identify risk factors at district level in Botswana using the logistic regression method. The results might help policy makers in the allocation of HIV/AIDS intervention resources and to decide on the factors to target first.

5.3. Contribution of the study

The main contribution of the study to the body of knowledge is in applying and comparing of different interpolation methods to investigate spatial distribution of HIV/AIDS in selected districts of Botswana. In addition, this study applied for the first time LISA and Getis Ord G_{i*} to BAIS data to investigate spatial patterns of HIV/AIDS in Botswana. The study has compared spatial maps from the current study with maps from previous studies. It has utilised Kulldorff scan and Getis Ord General G statistics to identify hot/cold spots of HIV/AIDS. The study has successfully unveiled risk factors that contribute to high HIV prevalence in some districts. Finally the study also contributes to the body of knowledge by using GIS which is limited in Botswana.

5.4. Limitations

This study's main limitation was the use of BIAS IV data which is cross-sectional and limits the analysis to the variables collected by the researcher. The study was limited to districts in Botswana that had a population greater than five thousands. Methods of determining the best spatial interpolation technique was limited to RMSE and variance. ArcGIS 10.1 calculates only one error (RMSE) for IDW and NN. Another limitation for this study was the inability to verify and countercheck the adequacy of the original data for inconsistencies and reliability. The researcher is not a GIS specialist.

5.5. Recommendations

This section puts in place recommendations for future studies on the basis of the findings of this study.

Recommendation 1

This study recommends that the government of Botswana should put in place programs that are suitable for each district. Prevention methods should be strengthened and improved to combat risk factors identified in the study.

The study identified alcohol abuse as the major risk factor associated with HIV/AIDS in all the districts. There is a need for intervention to target alcohol abuse in all the selected districts of Botswana, and to educate people on how alcohol abuse increases the risk of HIV/AIDS. The study recommends stiffer measures on alcohol use other than the current alcohol levy to combat alcohol abuse.

Recommendation 2

This study precisely identified hotspots and cold spots in the selected study area using LISA, Kulldorff's Scan and Getis-Ord Gi* statistics. The study recommends that HIV/AIDS programs, intervention and resources be targeted to the identified hotspots and their surroundings. Further research in this area should include other districts not included in this study so as to have current information about hotspots for the entire country.

Recommendation 3

The study accurately mapped the spatial distribution of HIV/AIDS in selected districts of Botswana using three interpolation methods. Additional analysis can be done using other deterministic and geostatistical interpolation techniques. The results of the two techniques can be compared to determine the best between the two, and may also be compared with the current study. The findings may be added to the existing literature on spatial distribution.

Recommendation 4

Spatial autocorrelation was employed to identify spatial patterns of HIV/AIDS at district level in Botswana and the result indicated that there is positive spatial autocorrelation. The finding might help policy makers and government to monitor if intervention exhibits similar patterns since near objects are more similar than distant objects according to Tobler's first law of geography (Zulu et al., 2014; Kalipeni and Zulu, 2008). Knowledge of spatial patterns might help the government of Botswana to restructure its intervention programs and come out with specific programs for each district.

More attention should be directed towards Central-Serowe, which this study identified as the hot spot for HIV/AIDS. Intervention should be re-enforced on education, age, alcohol use and condom use. Individuals with some education have no knowledge of the effects of HIV/AIDS hence there is need to educate them on the dangers HIV/AIDS. Many researchers have carried studies focused on interventions aimed at young people in the age group 15-29 years and very little on the older age groups. It is for this reason that this study recommends intervention programs for age groups 30-34, 35-39, 40-44.

Recommendation 5

The study recommends the use of LISA, Getis-Ord Gi*, Kulldorff scan, Moran's I and Getis-Ord General G statistics to be used to identify global and local patterns in other diseases like TB, Malaria and others. Also the study recommends the use of Geary's C statistics which is more sensitive to differences in small neighborhoods in analyzing autocorrelation.

Recommendation 6

This is the first study to compare current spatial maps with previous maps. This might help policy makers and health workers to determine if intervention has been effective and will present foundation to understand better the geographic background of activities.

5.6. Summary

The study aimed at analysing spatial distribution of HIV/AIDS using difference spatial techniques and comparing their effectiveness using 2013 BAIS IV data at district level in Botswana. The findings of the study using Moran's Index and Getis-Ord General G statistics has shown that HIV/AIDS in Botswana is spatially distributed and that neighbouring districts have similar HIV/AIDS characteristics. HIV prevalence in the selected districts was highest in Selebi-Phikwe and Francistown and least in Ngamiland South. The prevalence rate in Selebi-Phikwe is almost equivalent to the national prevalence. Hotspot detection tools have identified Central-Serowe as the epicentre (hotspot) of HIV/AIDS while Ngwaketswe, Kgalagadi North and south were identified as hotspots.

Ordinary kriging was found to be the best interpolation method for the spatial distribution of HIV/AIDS. The study identified risk factors that were associated with the spatial distribution of HIV/AIDS in selected districts. Alcohol abuse was identified as the main risk factor for all the selected districts. In general, the results indicated that HIV/AIDS is spatial distributed and that it is intense in the northeast part of the country.

REFERENCES

Adebayo, S. B., Fahrmeir, L. & Klasen, S. 2004. Analysing infant mortality with geo-additive categorical regression models: a case study for Nigeria. *Economic Human Biology*, (2):229–244.

Agyei-Mensah, S. 2005. The HIV/AIDS epidemic in sub-Saharan Africa: Homogeneity or heterogeneity? *Norwegian Journal of Geography,* 59: 14-25.

AIDS in Africa. 2013. Statistics and facts about HIV in Africa, Mewiser. www.medwiser.org/hiv-aids/around-the-world/. Date assessed: 4 September, 2015.

Airhihenbuwa, C.O. & Webster, J.D. 2004. Culture and African contexts of HIV/AIDS prevention, care and support. *SAHARA-J: Journal of Social Aspects of HIV/AIDS*, 1(1): 4-13.

Ansari, S. & Kale, K. 2014. Methods for Crime Analysis Using GIS. *International Journal of Scientific & Engineering Research*, 5(12):1330-1336.

Anselin, L. 1995. Local indicators of spatial association (LISA). *Geographical analysis*, 27(2):93-115.

ArcGIS Desktop 9.3 Help. 2011. *Environmental Systems.*

Asa, E., Saafi, M., Membah, J. & Billa, A. 2012. Comparison of linear and nonlinear kriging methods for characterisation of soil data. Journal of Computing in Civil Engineering 26(1):11-18.

Auvert, B., Taljaard, D., Lagarde, E., Sobngwi-Tambekou, J., Sitta, R. & Puren, A. 2005. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. Public Library of Science *2*(11): 298.

Avert. 2006. Averting HIV and AIDS: International AID Charity. www.avert.org Date assessed 21 June 2015.

Avert. 2014. Averting HIV and AIDS: International AID Charity. www.avert.org Date assessed 7 July 2015.

Avert. 2016. Averting HIV and AIDS: International AID Charity. www.avert.org Date assessed 14 July 2015.

Ayele, D.G., Zewotir, T.T. & Mwambi, H.G. 2012. Prevalence and risk factors of malaria in Ethiopia. *Malaria Journal,* 11(1):1.

Ayiga, N. & Letamo, G. 2011. Impact of male circumcision on HIV risk compensation through the impediment of condom use in Botswana. *African health sciences,* 11(4):550-559.

Azpurua, M.A. & Ramos, K.D. 2010. A comparison of spatial interpolation methods for estimation of average electromagnetic field magnitude. *Progress in Electromagnetics Research (*14):135-145.

Babu, B.S. 2016. Comparative study on the spatial interpolation techniques in GIS. International Journal of Scientific & Engineering Research, 7(2):550-554.

Bailey, R.C., Moses, S., Parker, C.B., Agot, K., Maclean, I., Krieger, J.N., Williams, C.F., Campbell, R.T. & Ndinya-Achola, J.O. 2007. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *The lancet, 369*(9562):.643-656.

Balk, D., Pullum, T., Storeygard, A. Greenwell, F. & Neuman, M. 2004. A spatial analysis of childhood mortality in West Africa. *Population, Space and Place,* 10(3):175-216.

Barankanira, E., Molinari, N., Niyongabo, T. & Laurent, C. 2016. Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention. *Biomedical Central public health, 16*(1):1.

Barnighausen, T. & Tanser, F. 2009. Rethinking the role of the local community in HIV epidemic spread in sub-Saharan Africa: a proximate-determinants approach. *HIV therapy,* 3(5):435-445.

Barrett, H.R. 2007. Too little, too late: Responses to the HIV/AIDS epidemics in sub-Saharan Africa. *Geography*, 87-96.

Beck, L.R., Kitron, U.R.I.E.L. & Bobo, M. 2002. Remote sensing, GIS, and spatial statistics: Powerful tools for landscape epidemiology. *Health Impacts of Global Environmental Change: Concepts and Methods, Martens P (edition). Cambridge University Press, Cambridge, New York, NY.*

Berkely, S. 1991. Parenteral transmission of HIV in Africa. *AIDS* 5(1):S87-S97.

Bhunia, G.S., Kesari, S., Chatterjee, N., Kumar, V & Das, P. 2013. Spatial and temporal variation and hotspot detection of kala-azar disease in Vaishali district (Bihar), India. *Biomedical Central Infectious Disease*, (13):64.

Blackstone, A. 2012. Principles of Sociological Inquiry–Qualitative and Quantitative Methods.

Boerma, J.T. & Weir, S.S. 2005. Integrating demographic and epidemiological approaches to research on HIV/AIDS: the proximate-determinants framework. *The journal of infectious diseases (*1):S61-S67.

Bohling, G., 2005. Introduction to geostatistics and variogram analysis. *Kansas geological survey*,1-20.

Boissonnat, J.D. & Cazals, F. 2000. Smooth surface reconstruction via natural neighbour interpolation of distance functions. In *Proceedings of the sixteenth annual symposium on Computational geometry* (pp. 223-232). ACM.

Borges, D.P., Franke, J., da Anunciação, Y.M.T., Weiss, H. & Bernhofer, C. 2016. Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theoretical and applied climatology*, *123*(1-2):335-348.

Botswana National Atlas, Aspects of Botswana's physical and human geography. http://www.atlas.gov.bw/html/chapA.html

Boulos, M.N.K. 2004. Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom. *International Journal of Health Geographics*, 3(1):1.

Briggs Henan University. 2010. Spatial statistics. www.utdallas.edu/~briggs/henan/10SAglobal.ppt

Bryman, A. & Bell, E. 2015. Business research methods, Oxford University press.

Burns N. & Grove S.K. 2005. The Practice of Nursing Research: Conduct, critique and utilization (5th Ed.). St. Louis Elsevier Saunders.

Burrough, P. & McDonnell, R. 1998. Principles of Geographical Information Systems. Oxford University Press, 132 – 161.

Caruso, C. & Quarta, F. 1998. Interpolation methods comparison. Computers mathematical application, 35(12):109-126.

Central Statistics Office 2013. Botswana Aids Impact Survey IV summary report, Central Statistics Office.

Central Statistics Office (CSO). 2013. Botswana AIDS Impact Survey IV, preliminary findings. Ministry of Health; Gaborone, Botswana: 2013.

Central statistics Office 2009. 2008 Botswana Aids Impact Survey III Statistical Report, Central Statistics Office.

Chai, H., Cheng, W., Zhou, C., Chen, X., Ma, X. & Zhao, S. 2011. Analysis and comparison of spatial interpolation methods for temperature data in Xinjiang Uygur Autonomous Region, China. *Natural Science*, *3*(12): 999.

Chilisa, B. & Kawulich, B.B. 2012. Selecting a research approach: paradigm, methodology and methods.

Chimoyi, L.A. & Musenge, E. 2014. Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011. *Biomedical Central Public Health*, 14:55.

Cohen, M.S. & Galvin, S.R. 2004. The role of sexually transmitted diseases in HIV transmission. *Nature Reviews Microbiology*, *2*(1); 33-42.

Cressie, N.A.C. 1993. Statistics for Spatial Data. Wiley, New York.

Daniel, M.L. 2000. The Demographic Impact of HIV/AIDS in Sub-Sahran Africa. *Geography: Journal of the Geographical Association*, *85*(1: 46

De Smith, M.J., Goodchild, M.F. & Longley, P. 2007. Geospatial analysis: a comprehensive guide to principles, techniques and software tools. Troubador Publishing Ltd.

Dean, H.D., Fleming, P.L. & Marconi, K. 2005. Uses of HIV and other public health data for HIV prevention and care planning. *AIDS Education and Prevention*, *17:*1.

Djukpen, R.O. 2012. Mapping the HIV/AIDS epidemic in Nigeria using exploratory spatial data analysis. *Geography Journal* 77(4): 555-569.

Eaton, L., Flisher, A.J. & Aarø, L.E. 2003. Unsafe sexual behaviour in South African youth. *Social science & medicine*, *56*(1):149-165.

Essex, M., Mboup, S., Kanki, P.J., Marlink, R.G. & Tlou, S.D. 2007. AIDS in Africa. Springer Science and Business Media.

Farahani, M., Vable, A., Lebelonyane, R., Seipone, K., Anderson, M., Avalos, A., Chadborn, T., Tilahun, H., Roumis, D., Moeti, T. & Musuka, G. 2014. Outcomes of the Botswana national HIV/AIDS treatment programme from 2002 to 2010: a longitudinal analysis. *The Lancet Global Health*, 2(1):e44-e50.

Ferguson, A.G. & Morris, C.N. 2007. Mapping transactional sex on the Northern Corridor highway in Kenya. Health and place, 13(2):504-519

Fotheringham, A.S., Bunsdon, C. & Charlton, M. 2002. Geographically Weighted Regression: the Analysis of Spatially Varying Relationships, John Wiley & Sons, Incorporated.

Fulchera, C. & Kaukinen, C. 2005. Mapping and visualizing the location HIV service providers: An exploratory spatial analysis of Toronto neighbourhoods. *Psychological and Socio-medical Aspects of AIDS/HIV*, 17 (3): 386-396.

Fytianos, K. & Christophoridis, C. 2004. Nitrate, arsenic and chloride pollution of drinking water in Northern Greece. Elaboration by applying GIS. *Environmental Monitoring and Assessment*, *93*(1-3): 55-67.

Gemperli A. & Vounatsou, P. 2003. Fitting generalised linear mixed models for point-referenced spatial data. *Journal of Modern Applied Statistics Method* (2): 497–511.

Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M. & Lengeler C. 2004. Spatial patterns of infant mortality: effects of malaria endemicity. *American Journal of Epidemiology*, 159: 64–72.

Getis, A. & Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. *Geographical analysis* 24 (3):189-206.

Glass, T.A. 2000. Update: spatial aspects of epidemiology: the interface with medical geography. *Epidemiologic Reviews*, 22(1): 136-139.

Glavanakov, S., White, D.J., Caraco, T., Lapenis, A., Robinson, G.R., Szymanski, B K. & Maniatty, W.A. 2001. Lyme disease in New York State: spatial pattern at a regional scale. *The American journal of tropical medicine and hygiene*, *65*(5):538-545.

Gould, P. 1993. The slow plague: Geography of the AIDS pandemic Cambridge: Blackwell Publishers.

Graham, A.J., Atkinson, P.M. & Danson, F.M. 2004. Spatial analysis for epidemiology. *American journal of Tropical Medicine and Hygiene*, 91(3): 219-225.

Gray, R H., Kigozi, G., Serwadda, D., Makumbi, F., Watya, S., Nalugoda, F., Kiwanuka, N., Moulton, L.H., Chaudhary, M. A., Chen, M.Z. & Sewankambo, N.K. 2007. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *The Lancet*, *369*(9562): 657-666.

Greener, R., Jefferis, K. & Siphambe, H. 2000. The impact of HIV/AIDS on poverty and inequality in Botswana. *South African Journal of Economics*, 68(5):393-404.

Gregory, I.N. 2003. Arts and Humanities Data Service (England). History Data Service. *A place in history: A guide to using GIS in historical research*. Oxford: Oxbow.

Grix, J. 2002. Introducing students to the generic terminology of social research. *Politics, 22*(3): 175-186

Guarneri, J.C. & Weih Jr, R.C. 2012. Comparing Methods for Interpolation to Improve Raster Digital Elevation Models. *Journal of the Arkansas Academy of Science, 66*: 77-81.

Guba, E.G. & Lincoln, Y.S. 1 994. Competing paradigms in qualitative research. *Handbook of qualitative research, 2*(163-194):105.

Guo, Q., Li, W., Yu, H. and Alvarez, O. 2010. Effects of topographic variability and lidar sampling density on several DEM interpolation methods. *Photogrammetric Engineering & Remote Sensing, 76*(6):701-712

Hajizadeh, M., Sia, D., Heymann, S.J. & Nandi, A. 2014. Socioeconomic inequalities in HIV/AIDS prevalence in sub-Saharan African countries: evidence from the Demographic Health Surveys. *International journal for equity in health, 13*(1):1.

Hanchette, C. L. 1999. GIS and Decision Making for Public Health Agencies: Child Lead Poisoning and Welfare Reform. *Journal Public Health Management Practice*, 5(4):41-47.

Hart, M.A. 2010. Indigenous worldviews, knowledge, and research: The development of an indigenous research paradigm.

Heritage, G. L., Milan, D. J., Large, A. R. & Fuller, I. C. 2009. Influence of survey strategy and interpolation model on digital elevation models quality. *Geomorphology*, 112(3):334-344.

Hixson, B.A., Omer, S.B., Del Rio, C. & Frew, P.M. 2011. Spatial clustering of HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations. *Journal of urban health, 88*(1):.129-141.

Hladik, F. & McElrath, M.J. 2008. Setting the stage: host invasion by HIV. Nature Reviews Immunology, 8(6):447-457.

Hofierka, J., Cebecauer, T. & Šúri, M. 2007. Optimisation of interpolation parameters using cross-validation. In Digital Terrain Modelling: 67-82. Springer Berlin Hiedelberg.

Hoshi, T., Fuji, Y., Nzou, S.M., Tanigawa, C., Kiche, I., Mwau, M., Mwangi, A.W., Karama, M., Hirayama, K., Goto, K. & Kaneko, S. 2016. Spatial distributions of HIV infection in an endemic area of western Kenya: guiding information for localized HIV control and prevention. *PloS one*, *11*(2):.e0148636.

Hsueh, Y.H., Lee, J. & Beltz, L. 2012. Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008. *Applied Geography*, 34:587-594.

Iacozza, J. & Barber, D. 1999. An Examination of the Distribution of Snow on Sea-Ice. Atmosphere-Ocean, 37(1):.21 – 51.

Ikeda, A.A. 2009. Reflections on qualitative research in business. *Revista de Gestão*, 16(3):49.

Iqbal, Z. and Zorn, C. 2010. Violent conflict and the spread of HIV/AIDS in Africa. *The Journal of Politics*, *72*(1):149-162.

Islam, T. & Tsuiki, M. 2016. Spatial Distribution of Air Dose Rate in Grazing Grassland. *Journal of Data Science*, *14*(1), pp.133-148.

Jeefoo, P. 2012. Spatial patterns analysis and hotspots of HIV/AIDS in Phayao Province, Thailand. *Archives des Sciences Journal*, 65(37): 37- 50.

Johnston, VerHoef, J.M., Krivoruchko, K. & Lucas, N. 2001. Using ArcGIS Geostatistical Analyst. ESRI.

Joint United Nations Programme on HIV/AIDS (UNAIDS), Global Report: UNAIDS report on the global AIDS epidemic 2013, 2013

Jongsthapongpanth, A. & Bagchi-Sen, S. 2010. Spatial and sex differences in AIDS mortality in Chiang Rai, Thailand. *Health and Place,* 16(4): 1084-1093.

Joshua, V., Bhoopathi, K., Paranjape, R. S., Subramanian, T. & Mehendale, S. 2014. Cluster detection of HIV infection for clients of female sex workers in selected districts of Tamil Nadu, South India. *Biomedical Central Infectious disease,* 14:027.

Kaiser family foundation. 2015. The global HIV/AIDS epidemic- fact sheet.files.kff.org/attachment/

Kalipeni, E. & Zulu, L. 2008. Using GIS to model and forecast HIV/AIDS rates in Africa, 1986 – 2010. *Association of American Geographers* 60(1): 33 -35.

Kandala, N. B., Campbell, E. K., Rakgoasi, S. D., Segwagwe, B.C. M. & Fako, T.T. 2012. The geography of HIV/AIDS prevalence rates in Botswana. *HIV AIDS (Auckland*), 4: 95.

Kandala, N.B. 2006. Bayesian geo-additive modelling of childhood morbidity in Malawi. *Applied Stochastic Model for Business Industry*, 22(2):139–154.

Kandala, N.B., Ji, C., Stallard, N., Stranges, S. & Cappuccio, F.P. 2007. Spatial analysis of risk factors for childhood morbidity in Nigeria. The American journal of tropical medicine and hygiene, 77(4):770-779.

Kandwal, R., Garg, P.K. & Garg, R.D. 2009 Health GIS and HIV/AIDS studies: Perspective and retrospective. *Journal of Biomedical Informatics*, (42): 748-55.

Kazembe, L.N. & Namangale J.J. 2007. A Bayesian multinomial model to analyse spatial patterns of childhood co-morbidity in Malawi. *European Journal of Epidemiology* (22): 545–556.

Kitron, U., Clennon, J.A., Cecere, M.C., Gürtler, R.E., King, C.H. & Vazquez-Prokopec, G. 2006. Upscale or downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. *Geospatial health, 1*(1):49.

Koff, W.C. & Berkley, S.F. 2010. The renaissance in HIV vaccine development—future directions. New England *Journal of Medicine*, 363(5):7.

Kozlov, A.P Heimer, R., Barbour, R., Shabolts, A., V., Hoffman, I. F. 2006. Spatial distribution of HIV prevalence and incidence among injection drug users in St Petersburg: Implication for HIV transmission. *AIDS*, 22(1):123-130.

Kristian, L. & Jason, G. 2005. Mapping the evolution of 'food deserts' in a Canadian city: Supermarket accessibility in London, Ontario, 1961–2005 *International Journal of Health Geographic,* 7(1):1.

Kulldorff, M. & Nagarwalla, N. 1995. Spatial disease clusters: disease and inference, *Journal of statistics in Medicine*, 14:799-810.

Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R. & Mostashari, F. 2005. A space-time permutation scans statistics for disease outbreak detection. *Public Library of Science Medicine*, 2:59

Kurtovic, S., Talovic, S. and Dacic, L. 2015. Global Competitiveness Indicators as a Determinant of Brownfield Investments in Serbia: Generalized Linear Model. *Research in World Economy, 6*(3):1.

Labaree, R. V. 2009. The methodology-Organisation of your social science research. libguides.usc.edu/writingguide/methodology.

Lam, N.S.N., Fan, M. & Liu, K.B. 1996. Spatial-Temporal Spread of the AIDS Epidemic, 1982–1990: A Correlogram Analysis of Four Regions of the United States. *Geographical Analysis*, *28*(2): 93-107.

LaScala, E. A., Gruenewald, P. J. & Johnson, F.W. 2004. An ecological study of the locations of schools and child pedestrian injury collisions .Accident Analysis and Prevention. 36(4):569-576.

Ledoux, H. & Gold, C. 2005. An efficient natural neighbour interpolation algorithm for geoscientific modelling. In *Developments in spatial data handling* 97-108 Springer Berlin Heidelberg.

Letamo, G. 2003. Prevalence of, and factors associated with, HIV/AIDS-related stigma and discriminatory attitudes in Botswana. *Journal of Health, Population and Nutrition*. 347-357.

Letamo, G. 2011. Misconceptions about HIV prevention and transmission in Botswana. *African Journal of AIDS Research* 6(2): 193-198.

Levin, D. M. 1988. The opening of vision: Nihilism and post-modern situation. London: Routledge.

Li, J. & Heap, A.D. 2008. A review of spatial interpolation methods for environmental scientists.

Li, X., Lee, S.L., Wong, S.C., Shi, W. & Thornton, I. 2004. The study of metal contamination in urban soils of Hong Kong using a GIS-based approach. *Environmental Pollution*, *129*(1); 113-124

Logan, J.R. 2012. Making a place for space: Spatial thinking in social science. *Annual review of sociology, 38.*

Lopman, B., Nyamukapa, C., Mushati, P., Mupambireyi, Z., Mason, P., Garnett, G. P. & Gregson, S. 2008. HIV incidence in 3 years of follow-up of a Zimbabwe cohort- 1998-2000 to 2001- 2003: contributions of proximate and underlying determinants to transmission. *International journal of Epidemiology* (37): 88-105.

Lu, G.Y. and Wong, D.W. 2008. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences, 34*(9), pp.1044-1055.

Luo, W., Taylor, M. C. & Parker, S. R. 2008. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *International Journal of Climatology, 28*(7):947-959.

Lwanga, S. K., Tye, C. & Ayeni, O. 1999. Teaching health statistics, Lesson and Seminar Outlines. Second edition. Singapore: World Health Organisation.

Madlala, L. S. 2008. Age-disparate and intergenerational sex in southern Africa: the dynamics of hyper vulnerability. *Aids*, 22:S17-S25.

Magadi, M. & Desta, M. 2011. A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: evidence from the DHS. *Health & place, 17*(5): 1067-1083.

Mair, A. & Fares, A. 2011. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of hydrologic engineering* 16(4):371-383.

Mala, S. & Sengupta, R. 2013. Geo-visual Approach for Spatial Scan Statistics: An Analysis of Dengue Fever Outbreaks in Delhi. *International Journal of Advanced Computer Science and Applications 4*(10)

Matshe, I. & Pimhidzai, O. 2008. Macroeconomic impact of HIV and AIDS on the Zimbabwean economy: a human capital approach. *Botswana Journal of Economics*, *5*(7): 32-45.

Mavedzenge, S.N., Olson, R., Doyle, A.M., Changalucha, J. & Ross, D.A. 2011. The epidemiology of HIV among young people in sub-Saharan Africa: know your local epidemic and its implications for prevention. *Journal of adolescent health*, *49*(6), pp.559-567.

McCoskey, S.K. 2003. A Spatial Correlation-Based Study of the Incidence of HIV-AIDS in Sub-Saharan Africa. Paper presented to the Eighth Annual Conference on Econometric Modelling in for Africa, Matieland, South Africa, 1–4 July.

Merrigan, M.B., Tafuma, T.A., Okui, L.A., Lebelonyane, R., Bolebantswe, J.M., Makhaola, K., Mine, M., Adhikary, R. & Chabikuli, N.O. 2015. HIV prevalence and risk behaviours among female sex workers in Botswana: results from the 2012 HIV/STI bio-behavioural study. *AIDS and Behaviour*, *19*(5): 899-908.

Messina, J. P., Emcha, M., Muwonga, J., Mwandagalirwa, K., Edidi, S. B., Mama, N., S. B., Mama, N., Okenge, A. & Meshnick, S. R. 2010. Spatial and socio-behavioural patterns of HIV prevalence in the Democratic Republic of Congo. *Social Sciences Medical Journal*, 71(8):1428–1435.

Ministry of Health, 2014. A strategy for changing the health sector. www.moh.gov.bw/Publications/policies/Botswana%20IHSP%20Final%20HLSP.pd

Moise, I. K. & Kalipeni, E. 2010. Application of geospatial analysis to surveillance data: A spatial examination of HIV/AIDS prevalence in Zambia. *Geography Journal*, 77(4): 525-540.

Moise, I. K., Cunningham, M. & Inglis, A. 2015. A process guide to monitoring and evaluation for informed decision making. Geospatial analysis in global health Measure and Evaluation.

Molatole, R. & Thaga, S.L. 2006. Interventions against HIV/AIDS in the Botswana Defence Force. *The Enemy Within: Southern African Militaries' Quarter-Century Battle with HIV and AIDS*, 19.

Moodley, Y. 2010. The use of GIS/GPS technologies in HIV/AIDS research. *South African Journal of Epidemiology and Infection*, 25(2), 33-34.

Moonan, P., Bayona, M., Quitugua, T., Oppong, J., Dunbar, D., Jost, K. J., Burgess, G., Singh, K. & Weis, S. 2004. GIS technology to identify areas of tuberculosis transmission and incidence. *International Journal of Health Geographies* 3:23.

Moran, P.A.P. 1948. The interpretation of statistical maps, *Journal of the Royal Statistics Society, Series B,* 10(2):243-251.

Morison, L. 2001. The global epidemiology of HIV/AIDS. *British Medical Bulletin*, 58(1):7-18.

Murphy, R.R., Curriero, F.C. & Ball, W.P. 2009. Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. *Journal of Environmental Engineering*, *136*(2):160-171.

National Geography Society, 2011 Geographic Information System http://nationalgeographic.org/encyclopedia/geographic-information-system-gis/

Nyindo, M. 2005. Complementary factors contributing to the rapid spread of HIV-I in sub-Saharan Africa: a review. *East African medical journal*, *82*(1).

Noor, A. M., Zurovac, D., Hay, S. I., Ochola, S. A. & Snow, R. W. 2003. Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Tropical Medicine & International Health*, 10(8), 917-926.

Nyblade, L., Field-Nguer, M.L., Coon, K., Ogden, J., Odolon, J., Obudi-Owor, A., Otim, C., Byakigga, J., Spebanja, P., Strickland, R.S. & Varga, C.A. 2001. Women, communities, and the prevention of mother-to-child transmission of HIV: issues and findings from community research in Botswana and Zambia.

Obidoa, C. A. & Cromely R. G. 2012. A geographical analysis of HIV/AIDS infection in Nigeria, 1991 – 2001*Journal of Social, Behavioural, and Health Sciences*, (6):13- 29.

O'Leary, E. S., Vena, J.E., Freudenheim, J.L. & Brasure J. 2004. Pesticide exposure and risk of breast cancer: a nested case-control study of residentially stable women living on Long Island. *Environmental Research*. (2):134-144.

Oliver, M.A. & Webster, R. 1990. Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System, 4(3):313-332.

Osei, F. B. & Duker, A. A. 2008. A Spatial and demographic patterns of Cholera in Ashanti region-Ghana. *International Journal of Health Geography*, 7:(1)1.

Otwombe, L. A. 2013. Spatial distribution and analysis of factors associated with HIV infection among Young People in Eastern Africa: Applied to the MEASURE Demographic and Health Survey data collected between 2007 and 2011.Johannesburg: University of Witwatersrand (Thesis- Masters).

Oyana, T.J., Rogerson, P. & Lwebuga-Mukasa, J.S. 2004. Geographic clustering of adult asthma hospitalization and residential exposure to pollution at a United States-Canada border crossing. *American journal of public health*, *94*(7):1250-1257.

Park, H. 2013. An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, *43*(2):154-164.

Peng, C.Y.J. & So, T.S.H. 2002. Logistic regression analysis and reporting: A primer. Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 1(1):31-70.

Peng, C.Y.J., Lee, K.L. and Ingersoll, G.M. 2002. An introduction to logistic regression analysis and reporting. *The journal of educational research*, *96*(1):.3-14.

Peng, Z., Cheng, Y., Reilly, K. H., Wang, L., Qin, Q., Ding, Z., Ding, G. ,Ding, K. Yu, R., Chen, F. & Wang, N. 2011. Spatial distribution of HIV/AIDS in Yunnan province, People's Republic of China. *Geospatial Health*, 5 (2):177-182.

Pfeiffer, D., Robinson, T.P., Stevenson, M., Stevens, K.B., Rogers, D.J. and Clements, A.C. 2008. *Spatial analysis in epidemiology* (Vol. 557976890). New York: Oxford University Press.

QGIS documentation spatial analysis interpolation
https://docs.qgis.org/2.2/en/docs/gentle_gis.../spatial_analysis_interpolation.h

Queensland Health. 2005. Report on GIS and public health spatial application. Public health services branch

Ramjee, G. & Wand, H. 2014. Geographical clustering of high risk sexual behaviours in "hot-spots" for HIV and sexually transmitted infections in Kwazulu-Natal. South Africa AIDS Behaviour, 18(2):317–322.

Riner, M. E., Cunningham, C. & Johnson, A. 2004. Public health education and practice using geographic information system technology. *Public Health Nursing*, 21 (1):57-65.

Rogerson, P. and Yamada, I. 2008. Statistical detection and surveillance of geographic clusters. CRC Press.

Rossman, G.B., & Rallis, S.F. 2003. Learning in the field: An introduction to qualitative research. 2nd edition, Thousand Oaks, 31-60.

Rubin, A. & Babbie, E.R. 2016. Empowerment Series: Research Methods for Social Work. Cengage Learning.

Ruiz, M.O., Tedesco, C., McTighe, T.J., Austin, C. & Kitron, U. 2004. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area. *International Journal of Health Geography*, 3: 8.

Saunders, M. N. & Bezzina, F. 2015. Reflections on conceptions of research methodology among management academics. *European Management Journal*, 33(5):297-304.

Saunders, M., Lewis, P. & Thornhill, A. 2009. Research Methods for Business students, fifth edition Rotolito.

Seloilwe, E.S. 2005. Factors that influence the spread of HIV/AIDS among students of the University of Botswana. *Journal of the Association of Nurses in AIDS Care*, *16*(3): 3-10.

Setianto, A. & Triandini, T. 2015. Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. *Journal of Applied Geology*, *5*(1).

Sharp, P.M & Hahn, B.H. 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine*, *1*(1), p.a006841.

Šiljeg, A., Lozić, S. & Šiljeg, S. 2015. A comparison of interpolation methods on the basis of data obtained from a bathymetric survey of Lake Vrana, Croatia. *Hydrology and Earth System Sciences*, *19*(8):3653-3666.

Simbayi, L., Shisana, O., Rehle, T., Onoya, D., Jooste, S., Zungu, N. & Zuma, K. 2014. South African national HIV prevalence, incidence and behaviour survey, 2012. Pretoria*: Human Sciences Research Council.*

Sipe, N.G. & Dale, P. 2003. Challenges in using geographic information systems (GIS) to understand and control malaria in Indonesia. *Malaria journal*, *2*(1):.36.

Small, E. 2009. Relationships and Social Influences as Predictors of Female Adolescent Condom Use. ProQuest.

Snow, J. 1855. On the Mode of Communication of Cholera, London, England, John Churchill.

Strickland, J. 2015. Predictive analytics using R. Lulu.com

Sunila, R. & Kollo, K. 2007. A comparison of geostatistics and fuzzy application for digital elevation model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*(2)

Tan, Q. & Xiao, X. 2014. Comparative analysis of spatial interpolation methods: an experimental study. Sensors & Transducers, 165(2):155.

Tanser, F. C. & Le Sueur, D. 2002. The application of geographical information systems to important public health problems in Africa. *International Journal of Health Geography*, 1(4).

Tanser, F., Barnighausen, T., Cooke, G. S. & Newell, M. L. 2009. Localised spatial clustering of HIV Infections in a widely disseminated Rural South African epidemic. *International Journal of Epidemiology*. 38(4) 1008-1016.

Tanser, F., Le Sueur, D. & Wilkinson, D. 2000. HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system. *Tropical Medicine and International Health* 5(1): 40-46

Taylor, S. J., Bogdan, R. & DeVault, M. 2015. Introduction to qualitative research methods: A guidebook and resource. John Wiley & Sons.

Tobler W.R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46(1) 234-240.

Uchudi, J., Magadi, M. & Mostazir, M. A. 2012. Multilevel analysis of the determinants of high-risk sexual behaviour in sub-Saharan Africa. *Journal of Biological Social Science,* 44(3):289–311.

UNAIDS 2004. UNAIDS report on the global AIDS epidemic. Geneva, Switzerland

UNAIDS 2006 Report on the global AIDS epidemic. Geneva, Switzerland

UNAIDS 2012. UNAIDS report on the global AIDS epidemic, 2012. Geneva, Switzerland.

UNAIDS global facts sheet HIV/AIDS 2014. Stepping Up the Pace. 20th International AIDS Conference. Melbourne, Australia.

UNAIDS. 2010. UNAIDS report on the global AIDS epidemic. The global reference book on the AIDS epidemic and response: http://www.unaids.org/globalreport. Date accessed 30 July 2015

UNAIDS. 2013 Global report. www.unaids.org/sites/default/files/media. Date assessed: 17 July, 2015.

UNAIDS. 2016. Global AIDS update. www.unaids.org/en/resources/documents/. Date accessed: 22 July 2015.

UNAIDS. 2015 Global AIDS response progress reportingwww.unaids.org/sites/default/files/media. Date accessed 27 July 2015

University of Bradford School of Management, 2013. Introduction to research and research and research methods. School of Management,1:1.

Van Houdt, R., Bruisten, S.M., Geskus, R.B., Bakker, M., Wolthers, K.C., Prins, M. & Coutinho, R.A. 2010. Ongoing transmission of a single hepatitis B virus strain among men having sex with men in Amsterdam. *Journal of viral hepatitis*, 17(2):108-114.

Waller, L.A. & Gotway, C.A. 2004 Applied spatial statistics for public health data. Wiley-Interscience, Hoboken.

Wand, H., Whitaker, C. & Ramjee, G. 2011. Geo-additive models to assess spatial variation of HIV infections among women in local communities of Durban. South Africa *International Journal of Health, 10*(28).

Webster, R., Oliver, M.A., Muir, K. R. & Mann, J. R. 1994. Kriging the local risk of a rare disease from a register of diagnoses. *Geographical analysis*, 26(2):168-185.

Weir, S. S., Morroni, C., Coetzee, N., Spencer, J. & Boerma, J. T. 2002. A pilot study of a rapid assessment method to identify places for AIDS prevention in Cape Town, South Africa. *Sexually Transmitted Infections*, 78(1).

Westercamp, N., Moses, S. Agot, K., Ndinya-Achola, J.O., Parker, C.Amolloh, K.O. & Bailey R.C. 2010. Spatial distribution and cluster analysis of sexual risk behaviours reported by young men in Kisumu, Kenya. *International Journal of health geography*, 9 (24)

William, B.G., Lloyd-Smith, J.O., Gouws, E., Hankins, C., Getz, W.M., Hargrove, J., De Zoysa, I., Dye, C. & Auvert, B. 2006. The potential impact of male circumcision on HIV in Sub-Saharan Africa. *PLoS Medicine*, *3*(7):262.

William, T. M. & Donnelly, J. P. 2001. Research Method and knowledge base.

Wong, D.W., Yuan, L. & Perlin, S.A. 2004. Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Analysis and Environmental Epidemiology* 14(5): 404-415.

World Bank, 2012. Prevalence of HIV, total (% of population ages 15-49). ***data.worldbank.org/indicator/SH.DYN.AIDS.ZS***

World Health Organization, 2004. The World health report: 2004: changing history. World Health Report 2004. Changing history

Yao, X., Fu, B., Lü, Y., Sun, F., Wang, S. & Liu, M. 2013. Comparison of four spatial interpolation methods for estimating soil moisture in a complex terrain catchment. *PloS one*, *8*(1):54660.

Yasrebi, J., Saffari, M., Fathi, H., Karimian, N., Moazallahi, M. & Gazni, R. 2009. Evaluation and comparison of ordinary kriging and inverse distance weighting methods for prediction of spatial variability of some soil chemical parameters. *Research Journal of Biological Sciences* 4 (1):93-102.

Ye, W., Hong, H.P. & Wang, J.F. 2014. Comparison of Spatial Interpolation Methods for Extreme Wind Speeds over Canada. *Journal of Computing in Civil Engineering*, *29*(6):401-4095.

Zandi, S., Ghobakhlou, A. & Sallis, P. 2011. Evaluation of spatial interpolation techniques for mapping soil pH. 19th International congress on Modelling and simulation, Perth, Australia, 12-16 December.

Zhang, Y., Xiao, Q., Zhou, L., Ma, D., Liu, L., Lu, R., Yi, D. & Yi, D. 2015. The AIDS epidemic and economic input impact factors in Chongqing, China, from 2006 to 2012: a spatial-temporal analysis. *British Medical Journal open*, *5*(3):006669.

Zhi-Hang, P., Yue-Jia, C., Reilly, K.H., Lu, W., Qian-Qian, Q., Zheng-Wei, D., Guo-Wei, D., Ke-Qin, D., Rong-Bin, Y., Feng, C. & Ning, W. 2011. Spatial distribution of HIV/AIDS in Yunnan province, People's Republic of China. Geospatial Health, 5(2):177-182.

Zhou, F., Guo, H.C., Ho, Y.S. & Wu, C.Z. 2007. Scientometric analysis of geo statistics using multivariate methods. *Scientometrics*, *73*(3):65-279.

Zhou, Y.B., Liang, S., Wang, Q.X., Gong, Y.H., Nie, S.J., Nan, L., Yang, A.H., Liao, Q., Song, X.X. & Jiang, Q.W. 2014. The geographic distribution patterns of HIV-, HCV-and co-infections among drug users in a national methadone maintenance treatment program in Southwest China. *Biology Medical Central infectious diseases*, *14*(1):1.

Zimmerman D, Pavlik C, Ruggles A. & Armstrong M.P. 1999. An experimental comparison of Ordinary and UK and Inverse Distance Weighting. Mathematical Geology 31: 375-390.

Zulu, L. C., Kalipeni, E. & Johannes, E. 2014. Analysing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994 – 2010. *Biology Medical Central Infectious Diseases* 14: 285