SOCIO-DEMOGRAPHIC FACTORS ASSOCIATED WITH HIV INFECTION: USING GPS IN HOME BASED HIV COUNSELING AND TESTING IN KALIMONI SUB-LOCATION, THIKA DISTRICT-KENYA

BY
Albino Luciani Mutanda

A thesis submitted to the University of Nairobi Institute Of Tropical and Infectious Diseases in partial fulfillment for a Degree of Master of Science in Medical Statistics
(MSc MedStat)

Institute of Tropical and Infectious Diseases
College of Health Sciences
University of Nairobi
©November 2009
Declaration

I, the undersigned declare that this project is my original work and has not been presented as a degree in any other university

Albino Luciani Mutanda
Signature ........................................ Date..................................................

The project has been submitted for examination with my approval as the University Supervisor

Dr. Thomas Achia,
School of mathematics,
University of Nairobi,
* P.O Box 30197-00100 Nairobi (Kenya)
Signature ........................................ Date..................................................

Mr. Lawrence Muthami (PhD c),
Center for Public Health Research,
P.O Box 20752-00202 Nairobi (Kenya)
Signature ........................................ Date..................................................
Dedication

I dedicate this work to my sons Telvin and Robert for their inspiration in my life and my wife June for her positive criticism.
Acknowledgement

I thank my family for their continuous prayers and support.

I am indebted to LVCT for providing me with data for my research project.

I greatly extend my appreciation to my Supervisors Dr. Achia and Mr. Muthami for their patience with me and guidance during the course of my studies and research.

I also acknowledge my teachers Mrs. Wang’ombe and Dr. Nguti for moulding me academically; UNITID staff: Prof Estambale-Director, Dr. Machoki-Deputy Director, Mr. Irungu-Senior Administrator, Olivia, Susan and Barasa for their tireless support.

All references have been clearly acknowledged at the end of this write-up
TABLE OF CONTENTS

Cover Page i
Declaration ii
Dedication iii
Acknowledgement iv
Table of Contents 1
Acronyms 3
List of Tables and Graphs 4
Abstract 5

1. Chapter One-Introduction 7
   1.1 Geography and Health 7
   1.2 Epidemiology of HIV in Kenya 8
   1.3 Government of Kenya HIV prevention strategies 13
   1.4 Motivating program-Home based HIV counseling and testing at LVCT 16
       1.4.1 Background 16
       1.4.2 Goal 16
       1.4.3 Rationale 16
       1.4.4 Design 17
           1.4.4.1 Study area 17
           1.4.4.2 Study design and sampling 18
           1.4.4.3 Methods 19
   1.5 Background to this project 23
       1.5.1 Hypothesis 23
       1.5.2 Research questions 24

2. Chapter Two-Literature Review 25
   2.1 GIS use in HIV intervention settings 25

3. Chapter Three-Methodology 30
   3.1 Spatial Data 30
   3.2 Types of spatial data 31
       3.2.1 Points 31
       3.2.2 Marks 31
       3.2.3 Covariates 32
   3.3 Exploratory Data Analysis 32
       3.3.1 Tests of complete spatial randomness 32
       3.3.2 The G-function: Nearest Neighbor Distance Distribution 35
       3.3.3 Geographically Weighted Regression 36
       3.3.4 Identifying outliers 38
3.4 Variable description 39
3.5 Spatial Analysis 40
3.6 Models 41
    3.6.1 Point level models 41
    3.6.2 Point process models 42
    3.6.3 spatial models using Akaike information criterion (AIC) 42

4. Chapter Four- Results and Discussion 44
    4.1 Descriptive and multivariate Logistic Regression Analysis of the risk 44
        Factors for HIV
    4.1a Adjusted Binary Logistic Regression Analysis of HIV risk 47
    4.2 Classification Decision Rule 47
    4.3 Point pattern of HIV Prevalence 47
    4.4 Geographically Weighted Regression (Inverse Distance Weighting) 48
    4.5 Spatial Autocorrelation 50
        4.5.1 HIV Outcome and Sex 50
        4.5.2 HIV Outcome and Marital Status 51
        4.5.3 HIV Outcome and Age 53
        4.5.4 HIV Outcome and Highest level of Education 54
        4.5.5 HIV Outcome and Occupation 55
    4.6 Model selection 56
    4.7 Discussion 56
    4.8 Ethics 57

5. Chapter Five: Conclusion 58

References 60
Appendix 1: Ordinary Least Squares 65
ACRONYMS

AIC: Akaike Information Criterion
AIDS: Acquired Immune Deficiency Syndrome
ART: Anti-Retroviral Therapy
BCC: Behavior Change Communication
CBS: Central Bureau of Statistics
CDC: Centers for Disease Control and Prevention
CORPS: Community Owned Resource Persons
CSR: Complete Spatial Randomness
DASCO: District AIDS & STIs Coordinator
GIS: Geographic Information Systems
GoK: Government of Kenya
GPS: Geographic Positioning System
GWR: Geographically Weighted Regression
HBHTC: Home Based HIV Testing and Counseling
HIV: Human Immuno-deficiency Virus
IMF: International Monetary Fund
KAIS: Kenya AIDS Indicator Survey
KDHS: Kenya Demographic and Health Survey
KMOT: Kenya Modes of Transmission Study
LVCT: Liverpool VCT, Care and Treatment
NACC: National AIDS Control Council
NCHSTP: National Center for HIV, STD, and TB Prevention
NHSSP: National Health Sector Strategic Plan
OR: Odds Ratio
PMTCT: Prevention of Mother to Child Transmission
SPSS: Statistical Package for Social Scientists
STI: Sexually Transmitted Infections
UNAIDS: Joint United Nations Program on HIV/AIDS
VCT: Voluntary Counseling and Testing
WHO: World Health Organization
Tables and Graphs

Figure 1: HIV Prevalence in Kenya by Province 10
Figure 2: HIV Prevalence by Age & Sex 12
Figure 3: Map of Thika District Sub-Locations – Kalimoni in colored 18
Figure 4: Point Pattern of HIV Prevalence 47
Figure 5: Geographically Weighted Regression (Inverse Distance Weighting) 48
Figure 6: HIV outcome and sex 50
Figure 7: Ripley’s K Function 51
Figure 8: HIV outcome and Marital Status 51
Figure 9: HIV outcome and Age groups 53
Figure 10: HIV outcome and Level of Education 54
Figure 11: HIV outcome and Occupation 55
Figure 12: Script ordinary Least Squares 65

Table 1: Descriptive and Multivariate Logistic Regression Analysis of the risk factors for HIV 45
Table 2: Adjusted Binary Logistic Regression Analysis of the risk factor for HIV 45
Table 3: Classification Decision Rule 47
ABSTRACT

Background: From the "Kenya Modes of Transmission" (KMOT) study (Haile G. 2008), Kenya has a mixed epidemic with (a) National prevalence ranging from 6.7% (KDHS2003) to 7.4% (KAIS 2007) 7.1% translated to approximately 1.33 million Kenyans aged 15-64 living with HIV in 2007 while casual heterosexual sex contributing two thirds new infections and (b) Great regional variations, ranging from almost 1% (Northeast Province) to above 12% (Nyanza Province, and up to 30% in some fishing communities of Districts adjacent to Lake Victoria area). The KMOT study recommended that more research on the behaviour and mapping of most-at-risk individual(s), cultural issues requiring behaviour change and uptake of HIV services be conducted to establish where, how and who gets the next new HIV infection. Such research would require the use of geospatial multilevel and multidimensional survey

Methods: All 13,060 participants who consented to be counseled and tested for HIV drawn from a population under continuous demographical surveillance were linked to their homesteads and geolocated using a geographical information system (accuracy of <2 m). Individuals and/or families were counseled and tested as per national counseling and testing guidelines. In addition to capturing their behavioral, social and demographic variables, spatial coordinates of their homesteads were recorded as well. Point patterns were used to show geospatial variations while spatial autocorrelations were used to produce robust estimates of HIV prevalence that varied across continuous geographical space and discreet demographic factors (sex, age, marital status, education and occupation). Neighborhood spatial effects with a threshold of 0.135 degrees (15 km radius) were also applied to identify clusters of prevalence (P < 0.05).

Results: The results reveal considerable geographical variation in local HIV prevalence (range = 0.001-2.16) within this relatively homogenous population and provide clear empirical evidence for the localized clustering of HIV infections. Four discreet spatial clusters with filled contours ranging from 0.02-0.97
were identified by using ordinary least squares autocorrelation after testing for residual autocorrelation $(P \leq 0.05)$, within the study area.

**Conclusions:** The findings show the existence of several clustered HIV prevalence of varying intensity contained within the study community. Despite the overall low prevalence of HIV in Kalimoni sub-location, the results support the need for interventions that target socio-geographic spaces (clustered villages) at greatest risk to supplement measures aimed at the general population.
Chapter One: Introduction

1.1 Geography and Health

Human geographers consider interactions between humans and their environment, the unique characteristics of specific places, and spatial patterns in human activities. As historians consider the importance of time in making sense of the world, so do geographers focus on the importance of space and place. The geographer in his work relies extensively on maps and in this respect there would seem to be a useful field for joint exploration between geographer, human ecologist, epidemiologist and medical statistician (Melvyn, 1959). The mapping of statistical data is now generally accepted to be the method of attacking many problems in human medicine and its obvious importance in the understanding of arial relationships of diseases is such that it is somewhat surprising that more is not being done by the medical profession. The comparison of endemic areas with areas free from the disease may well throw valuable insights on causation. Once the active cause and the predisposing associations of a disease are defined, the path is open to sound preventive measure. If the cause is unknown the difficult early stages of the problem may well be clarified by a map showing the geographical variations of the particular disease. The deeper causes of the distribution can then be studied and it may well be that the distributional patterns displayed are paralleled by those of some other known phenomenon. These, in turn, may give a clue to causation. This is not to suggest that problems are elucidated merely by mapping broad arial relationships. Rather, it is thought that useful pointers and relationships might well emerge from such an exercise.

The spread of HIV infection is often associated with geographic factors such as population mobility, accessibility and proximity to high transmission or urban areas and geographic
distribution of populations at greater risk of infection. The HIV epidemic varies substantially from one geographic area to another, and three broad epidemic categories describe the diversity of features observed globally: low epidemic settings, centralized epidemics and generalized epidemics (Karim et al. 2007). Maps of trucking and trade routes, male circumcision and HIV prevalence and population mobility in urban areas have shown that geographic factors are important in understanding the risk of HIV infection as consistently identified in the literature on HIV (Montana et al. 2007, Arroyo et al. 2006; Coffee et al. 2005; Tanser et al. 2000; Moses et al. 1990; Bongaarts et al. 1989). Understanding both physical and social environments is a vital to planning and implementation of community-based intervention methods. Even in HIV treatment, spatial disparities are still evident, for instance the introduction of highly active antiretroviral therapy in industrialized countries has transformed AIDS from an inevitably fatal condition to a chronic, treatable condition, but this goal has yet to be realized in most resource-constrained settings that bear a disproportionate burden of HIV infection (Karim et al. 2007).

1.2 Epidemiology of HIV in Kenya

Kenya continues to have a severe, generalized HIV epidemic, but in recent years, the country has experienced a notable decline in HIV prevalence, attributed in part to significant behavioral change and increased access to ART (Wikipedia Kenya Profile, 2008).

HIV/AIDS spread rapidly in Kenya during the 1990’s reaching prevalence rates of 20-30% in some areas of the county. The prevalence subsequently declined in some areas in Kenya but
remained stable in others. National prevalence declined significantly from a peak of 10% to under 7% in 2004. This trend is supported by data from national surveys which document changes in behavior towards fewer partners, less commercial sex, greater condom use and late age at sex debut. The Kenya Demographic and Health Survey (KDHS, 2003) revealed that 6.7% of adults tested are infected with HIV.
The HIV prevalence rates among both women and men in 2007 were higher than the rates observed in 2003. According to KAIS for both females and males, HIV is occurring in all age groups. There are, however, some differences in prevalence across the life span. Among youth aged 15-24, women are 4 times more likely to be infected than men (6.1% compared to 1.5%). A higher proportion of Kenyans aged 30-34 are currently infected with HIV than in any other age category. The decline in prevalence among women after age 34, and among men after age 44 could represent a decline in new infections in older age groups or an increase in HIV-related deaths in these age groups. The burden of infections is statistically higher among females than males until age 35 after which the ratio of male to female infections approaches 1 to 1.
The distribution of HIV infections varies greatly across Kenya. Prevalence remains highest in Nyanza at 15.3%, more than double the national prevalence estimate. Other provinces with rates similar to or higher than the national level are Nairobi (9.0%), Coast (7.9%), and Rift Valley (7.0%). Prevalence in Eastern is 4.7% and in Central, 3.8% of the adult population is infected. North Eastern province has the lowest adult HIV prevalence at 1% (KAIS, 2007).

The provincial estimates for HIV prevalence among 15-49 year olds in 2007 were similar (within 1%) to estimates from KDHS 2003 for Nairobi, Central, Eastern and Western Provinces. The Coast experienced a striking increase in the proportion of adults living with HIV; the proportion of HIV infected adults was 2.3% points higher in 2007 than in 2003, representing a 40% increase in HIV prevalence. Similarly in Rift Valley, the increase in HIV prevalence of 2.1% points represents a 40% increase since 2003. About three quarters of Kenyans live in rural areas of the country. Among those ages 15-64, 7% are infected with HIV. In urban areas, the prevalence is 9%. Because of different population sizes across provinces, prevalence estimates may not provide the complete picture of HIV burden in a province. Though the proportion of infected adults in the Coast and Nairobi is higher than the proportion in Rift Valley, the number of infected adults in Rift Valley (estimated 322,000) was greater than in Coast (estimated 135,000) or Nairobi (estimated 176,000). Together, Nyanza and Rift Valley are home to half of all HIV-infected adults.
Kenya's HIV/AIDS prevalence rate (the percent of people living with the disease) is just below that of the sub-Saharan African region overall (6.7% compared to 7.5%). Recent data indicate that the country's HIV prevalence rate may be on the decline in some areas. However, the HIV/AIDS epidemic poses significant challenges to low-income countries (UNAIDS 2008).

From the recent "Kenya Modes of Transmission" (MOT) study (Haile G. 2008), Kenya has a mixed epidemic with (a) National prevalence ranging from 6.7% (KDHS2003) to 7.4% (KAIS 2007) and casual heterosexual sex practices is contributing two thirds of new infections and (b) Great regional variations, ranging from almost 1% (Northeast Province) to above 12% (Nyanza Province, and up to 30% in some fishing communities of Districts adjacent to Lake Victoria.
area). The MOT study recommended that more research was required on the behaviour and mapping of the most-at-risk individual(s), cultural issues requiring behaviour change and uptake of HIV services be conducted to establish who gets the next HIV infection, how and where. Such a research design would require the use of geospatial multilevel and multidimensional survey.

1.3 Government of Kenya HIV Prevention Strategies

The Government of Kenya (GoK) first established the National AIDS Control Council (NACC) in 1999, and has a National Strategic Framework for HIV/AIDS for 2005-2010 (UNAIDS, 2008) with an aim of, among other things, providing clear and agreed vision, goal and targets for the national response to the fight against HIV. To ensure long-term sustainability of the HIV/AIDS programmes, the GoK linked the process of implementing KNASP to the Economic Recovery Strategy for Wealth and Employment Generation (2003-2007) and the Kenya Government budgetary cycle (NACC, 2005)

A review of the most recent HIV epidemiology data shows that heterosexual transmission remains the most prominent mode of transmission in all areas of Kenya. However, heterosexual transmission occurs in a variety of types of sexual encounters: between married couples or steady sexual partners, concurrent sexual partnerships (e.g. one person with both a steady, long-term partner and a casual partner, or one person with more than one steady partner), casual sexual partners, and a range of transaction-based sexual practices.

The Government of Kenya as stated in the Kenya National Strategic Plan for HIV/AIDS (KNASP) for 2005-2010 has prioritized HIV prevention by reducing the number of new infections through
decreasing the risk of infection in the general population and decreasing high risk behaviors which make particular groups vulnerable to HIV infection. The prevention strategies contained therein are informed by the achievements, weaknesses and lessons learnt from previous prevention programs. New evidence based approaches to the prevention of HIV infection will continue to be sought to ensure changes in the general population and vulnerable groups (NACC, 2005). Combinations of HIV prevention strategies have been given first priority.

The first prevention strategy is increasing availability and access of counseling and testing as a key sexual and behavior change strategy. This includes both the ‘traditional’ VCT approach and newer models like provider “Initiated Counseling and Testing” (PITC) and home based HIV counseling and testing (HBHTC). Individuals who test HIV negative are motivated to guard against their sero-status, while those who test HIV positive are counseled on how to protect their partners from infection and be referred to comprehensive care and treatment services. The priority is to scale up HIV counseling and testing and referral to appropriate services, through effective communication strategy to influence accelerated increase in the number of people tested for HIV in the facility setting (KNASP, 2005).

The second important prevention strategy is condom promotion, which is a key methodology of prevention of HIV and other sexually transmitted infections (STIs). This is carried out through promotion of correct and consistent condom use among the general population and those at higher risk of HIV infection and strengthening marketing programs to enhance availability and affordability of condoms for the high-risk locations. Again this is achieved through scaling up of
condom distribution throughout the county, supported by information, education and communication (IEC) strategy targeting vulnerable groups (KNASP, 2005).

The third strategy is implementing an effective targeted behavior change communication (BCC) strategy. This is through implementing a capacity building program to the health workers through training and provision of re-use prevention injection equipment and appropriate health waste management. The BCC strategy focuses on reducing demand for unnecessary injections among communities. In addition to injection safety, infection prevention and control policy is also promoted to serve as a means of reducing health worker exposure to HIV infection. These government strategies also aims at reducing the number of young people having sex by the age of 15, and promote abstinence and or consistent practice of safe sex among those who are most vulnerable and the general population (NACC, 2005)

Other strategies include expanding prevention of mother to child transmission (PMTCT) of HIV, strengthening of STI and HIV program linkages, improving availability of safe blood supplies and ensuring that prevention and treatment efforts do not lead to any increase in high risk behavior which could undermine prevention efforts (NACC, 2005)
1.4 Motivating Program-Home Based HIV counseling and Testing at LVCT

1.4.1 Background

The data analysed are part of an ongoing home-based counseling and testing program run by Liverpool VCT Care and Treatment (LVCT) in Thika district Kenya. LVCT is a Kenyan non-governmental organization involved both in the provision of HIV counseling and testing services and the use of operational research to feedback into policy and influence Kenya’s national response to HIV. LVCT has had a working relationship with Thika district’s health management team since 1999, when the first two VCT sites were operationalized in government health facilities.

HBHTC is a new innovative approach to counselling and testing aimed at enhancing community access to counselling and testing by transferring the VCT services from the facility to the household level. The HBHCT concept borrows a lot from the National Health Sector Strategic Plan (NHSSP II-2005-2010).

1.4.2 Goal

To determine the level of HIV/AIDS reduction after enhancing diversified community access to the HTC to household level targeting families and couples.

1.4.3 Rationale

The GoK has targeted 80% of Kenyans to know their HIV status by 2010. However, this requires a diversity of approaches to counseling and testing which acknowledges shifting paradigms, increased efforts towards prevention, strengthened coordination and changing and unknown
vulnerabilities. Successful models for community based services have been described in Uganda and other neighboring countries and this has formed the basis for the implementation of the same in Kenya (Matovu, 2002)

1.4.4 Design

1.4.4.1 Study Area

Despite its proximity to Nairobi (see figure 3), Thika is a rural district where people practice subsistence farming alongside larger commercial coffee, tea, pineapple, sisal and flower farms. It has an estimated population of 645,713 (Wikipedia, 2009). Data from antenatal and demographic health surveys revealed that HIV sero-prevalence rates were between 4-5% based on sentinel sites (UNAIDS 2006). The selected sub-location (Kalimoni) has an estimated population of 120,677 (CBS, 2000). It is divided by a main road that links Nairobi to Thika and the selected villages comprised of a mixture of localities that were close to road, far from the road, near an industry (a commercial sisal farm) and an informal settlement (slum). The majority of the population lived in scattered homesteads that are not concentrated into villages or compounds. The study area is typical of many rural Kenyan settings in that while it is largely rural it also contains formal urban township and a series of high-density settlements located predominantly along the major transport routes.

Figure 3 below shows the map of Thika district demarcated into sub-locations. The colored parts show the original boundaries of the wider sub-location before it was further subdivided. The study area thus comprises of homesteads drawn from the older Kalimoni as of 1999 and not the current smaller sub-location.
1.4.4.2 Study Design and Sampling

This was a cross-sectional prospective program that provided HTC services. Quota sampling—a non-probability equivalent of stratified sampling approach was used in the program implementation. Like stratified sampling, the villages (strata) were first identified and their proportions are as represented in the population noted. However, whole homesteads within the villages were sampled. Then convenience sampling (consent by household members) was used to select the required subjects.
1.4.4.3 Methods

The HBHTC program began piloting HCT using GPS systems in June 2007 in Kalimoni. Eight villages out of 33 were selected in consultation with the District Public Health Officer. Information about the new services was announced in villages in advance of starting the exercise. The team of community mobilizers gave information about the CT services and answered questions as part of baseline assessment. Trained HIV counseling and testing counselors visited homes on a door to door basis in the 8 villages over a six month period. Additional training was provided to counselors on aspects of home entry, on family planning, couple and family counseling and on use of GPS. Household members who requested testing received standard pre- and post-testing counseling with quality assured rapid whole blood testing. A choice of privacy in HIV testing individually or testing as a group was offered. All household members were offered testing regardless of age, with parental consent being required for testing of children and infants. Data on demographics, risk characteristics and where appropriate test outcomes and referrals were collected using a standard data collection form. Consent was obtained prior to testing, including consent to follow up positive clients who did not attend referral points. All 13,060 adult and children residents who consented to counseling and a HIV test were geo-located to their homestead of residence. The latitude, longitudinal and where applicable elevation and depression of client households were measured from a Geographical Positioning System GPS (Germain E-trex type) gadget. Each household and its individual members were given a unique GPS code acting as a primary identifier. The data was mapped, entered and analyzed using Epilinfo and ESRI GIS. Data collected during the same
period at the fixed VCT sites in the district were used for comparison. To achieve these, the following steps were followed:

a. Exploration

This was a key fact finding mission to enable the service providers coming into the community gain as much knowledge and understanding of the community’s situation as possible. It was done by carrying out self discovery, exploration, using LVCT’s 3Ls tool; “Look, Listen and Learn”. The findings from these explorations were written and shared with the community.

b. Protocol

This stage entailed identifying the gatekeepers—both formal and informal community leaders in order to enter the community through them to formalize the process. Respect for protocol and explaining the purpose of the HBHTC program thoroughly to the leaders was very important. They were involved from the initiation of the program within the community up to the end; this enhanced community ownership, increased service demand, guaranteed security to the service providers and freedom to work with the community.

c. Community meeting

This stage involved provider discussions with the key individuals at every control point down to the household level whilst explaining at every level the intended implementation processes, extent of involvement and how the community would benefit.
d. Participatory planning

Participatory planning was done by service providers, community own resource persons (CORPS) and MoH partners. The process involved identification of resources, capacity and existing gaps in order to set objectives for making desired results by identifying activities to be carried out, how they would be done, where and when. They would then assign responsibilities and resources with specific indicators to measure achievement of targets.

e. Community mobilization

Community mobilization was a continuous process that helped in creating of awareness and sensitizing individuals, families and community as a whole to enable them take action by use of print, audio, audiovisual and none print media at provincial administration gatherings like chief’s barazas or common places as places of worship and market areas.

f. Counseling and testing

Counseling and testing was done to individuals, couples and/or families who consented to services provided. Parental consent was sought from parents who were willing to have their children tested. Individuals who turned positive were given appropriate referral

g. Support supervision

The supervision of counselors was based on the counselors’ supervision model that included; clients well being, emotional well being of the counselor, capacity building of counselors with adequate knowledge and skills based on continuous improvement, management and administrative issues and development of support supervision strategy with a check list.
h. Monitoring

Monitoring tools and a system that captured all HBHCT indicators were developed. Data collection mechanisms were formulated. Data entry and analysis was done in Epi Info statistical software. Feedback and reporting was given to the district health management team (DHMT) and donors for programmatic improvement and informed decision making.

i. Evaluation

Periodic evaluation of the program was done to determine the level of achievements, challenges in implementation and areas for improvement and defining HBHCT model for replication nationwide.

j. Quality assurance

This was a systematic and planned approach to monitoring, assessing and improving counseling and testing services on a continuous basis. Quality assurance in HBHCT focused on meeting clients’ needs and expectation, ensuring standards were met and quality of services maintained, using data to inform if standards were met and encouraging multi disciplinary teams to enhance problem solving and quality improvement.

k. Results

HBHCT was identified as key to achieving counseling and testing targets due to the large numbers of clients who received CT services. The results also showed increased couple counselling and testing compared to VCT. There was increased disclosure to couples or sexual partners. Increased community participation and involvement was also noted which led to
decreased stigma and discrimination. Important lessons learnt by LVCT included importance of knowledge and skills in community strategies as key to implementation of HBHCT including motivation of CORPS. Re-orientation of facility based VCT counsellors to community based counsellors is was also a challenge. The biggest challenge was meeting demands for other services like malaria nets, water purifiers and general disease diagnosis by community members.

1.5 Background to this Project

Knowledge of local variation in levels of HIV prevalence is important for prioritization of areas for intervention and allocation of resources in small homogeneous and heterogeneous populations. This study applies exploratory and relational spatial analytical techniques to investigate the micro-geographical patterns and clustering of HIV prevalence by demographic factors in Kalimoni-Thika District.

1.5.1 Hypothesis

H₀ HIV prevalence in different regions within the study area is spatially independent.

Hₐ HIV prevalence is not spatially independent.
1.5.2 Research Questions

The study sought to determine any spatial relationships in clients who tested HIV positive from LVCT's home based HIV testing and counseling (HBHTC) program in Kalimon sub-location.

In particular it sought to:

1. Seek if there exist any spatial clusters in clients who test HIV positive.

2. Determine spatial autocorrelations between HIV positive clients and sex, marital status, age, highest level of education and occupation.

3. Develop a spatial model with factors that best predict HIV outcome.
Chapter Two: Literature Review

2.1 GIS use in HIV intervention Settings

Where people live affects their health, nutrition and access to health care services. Spatial epidemiology involves using GIS to locate and track the progression of disease in order to identify epidemiological trends (Rushton, 2003) and suggest appropriate interventions. For instance by identifying the water source responsible for an outbreak of cholera in London and mapping the locations of those afflicted in 1854, an English physician, John Snow, demonstrated how mapping can be used in epidemiological research (Osborne, 2009). GIS is a means of capturing, storing, and analyzing data, and using specialized software to plot it on maps or to perform advanced spatial statistical analysis. Public health practitioners use basic GIS applications to map the locations of health facilities or ART patients in a district. They can use the maps to analyze different pieces of data in relation to each other, forming and testing hypotheses about interventions that may improve services. For instance, if maps created using GIS showed most ART patients in a district lived far from facilities providing services, healthcare professionals would posit that they may need to improve patient access. Multiple maps created through GIS can be overlaid, making it easier to see relationships between different sets of data. More complex investigations using GIS data and its spatial analysis tools can predict the geographic spread of disease, demonstrate temporal disease trends, or analyze health service gaps.

GIS provides a holistic public health approach that promotes the well being of human populations through organizing data about who we are, where we live, and how we live within
a geographic framework. The health of human populations reflects the complex interplay between population characteristics and the environment. Genetic makeup can predispose certain populations to chronic or acute conditions. Cultural factors, such as stress, economic status, and access to health care, can play a significant part in disease onset. For example, heart disease, cancer, and alcoholism—leading causes of death in the United States—are produced by multiple, interrelated factors rather than a single infectious agent (Davenhill 2002).

GIS incorporates data that describes population characteristics, socioeconomic conditions and the landscape, and analyzes the spatial relationship of these factors. In addition to integrating and analyzing health related data, this technology promotes data sharing through the use of standard formats and a highly efficient communication tool—the map (Bolstad 2005). Mapping the geographical distribution of HIV prevalence among different population groups may assist in interpreting both the GoK national coverage of the HIV surveillance system as well in explaining differences in levels of prevalence. The UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, in collaboration with the WHO Public Health Mapping Team, Communicable Diseases, continues to produce maps showing the location and HIV prevalence in relation to population density, major urban areas and communication routes in Kenya (UNAIDS/WHO, 2004).

GIS has continued to be used in public health epidemiological studies for tracking sources of diseases and the movements of contagions to ensure effective responses to disease outbreaks by identifying at-risk populations and targeted interventions.
Forces not specific to public health are also encouraging the spread of GIS use. Data standardization and the development of the harmonized health system infrastructure like mapping health centers support the use of GIS, not only for public health, but for all government and economic applications. Geospatial boundary and demographic data is more readily available from government and commercial sources in a variety of formats now than before. In recent health researches in Africa, GPS technology has been increasingly used. Although GIS methodology forms part of the basis of many public health and epidemiological initiatives in the developed world (Bullen, Moon, & Jones, 1996), the developing world, especially Africa, lags far behind in the application of GIS to the study of health problems (Oppong, 2000), with the exception of malaria (Martin, Curtis, Fraser, & Sharp, 2002).

In recent decades, participatory research methods have emerged as valuable tools. There is a strong link between public health and location. For instance, geographic accessibility to health facilities is an important factor in ensuring patients receive necessary care. The spread of diseases, such as HIV and malaria, can be affected by geographic factors or spatial clustering. Geographic information systems (GIS) supported by a strong spatial data infrastructure and quality routine health data can give planners valuable information to address these issues of planning, monitoring and evaluation (M&E).

There is increased albeit low research on HIV prevalence in Africa utilizing Geographic Information Systems (GIS) technology to its full potential. For instance a study applying GIS analysis to factors related to HIV prevalence (Tanser and le Seuer 2002) estimated local-level HIV prevalence rates using data obtained from antenatal care providers, and found a
correlation between HIV prevalence and proximity of local households to a primary or secondary road. Another study used econometric techniques to estimate the spatial correlation of HIV infection across international boundaries (McCoskey 2003). GIS has also been used to locate individuals who are at risk for HIV in order to make decisions about prevention services. For example, information collected from sex workers in Kenya was used to create GIS maps that identified truck stops with high volume of transactional sex (Ferguson & Morris, 2007). However, geographic information systems have been used to estimate and analyze the spatial distribution of other communicable diseases, such as TB in the United States (Moonan et al. 2004).

GIS has been used to map the number of HIV services in a given area in order to understand access to prevention and health care (McLafferty, 2003). The US Centers for Disease Control and Prevention (CDC) led a major initiative between 2000 and 2002 to map all the community-based organizations providing CDC-funded prevention services (Hanchette, 2005). The CDC collected data about the location of these programs and their geographic service area(s). The resulting *HIV Prevention Services [GIS] Database* can be used to inform decisions about prevention services by describing their spatial relationships to each other and to high risk communities (Hanchette et al, 2005).

Montana (2007) while quoting Mishra et al (2006), notes that the increasing need for more precise data on the HIV epidemic has led to more population-based Demographic and Health Surveys (DHS) utilizing GIS since they provide representative estimates for both women and
men, for geographic regions, and by age groups. They also offer enhanced linkage of HIV status with individual respondent and household characteristics. The linked surveys allow for the analysis of geographical factors, behavior, knowledge and background characteristics as they relate to HIV status.
Chapter Three: Methods

3.1 Spatial Data

There are three components to spatial data: features, supports, and attributes.

A feature is an object with a specific spatial location and distinct properties. There are several types of spatial features:

Point: a precise location, $S$, in space; a dot on a map. For example, a point could be the geographic location of a house or the location of a landmark.

Line: a sequential collection of connected points for instance roads, rivers and geographical boundaries.

Area: a region enclosed by lines. Sub-locations, divisions, counties, states are all examples of areal spatial objects.

Volume: a three-dimensional object having height or depth (vertical extent) as well as horizontal extent e.g. geologic formations such as aquifers.

A collection of features of the same type is called a feature class. Each feature is of a certain size and shape and has a specific spatial orientation. Taken together, these properties form the support of the data. Points, or spatial locations, have the smallest support. They have zero size, no shape, and no orientation. Lines have length and can indicate direction. Regions have area and boundaries that may impose properties on the associated features. For example, a circle and a rectangle are both areal features, yet they are inherently different spatial objects even if they have the same area.
Attributes are observations or measured values associated with features (e.g. the number of schools in a sub-location). In any given analysis there may be a single type of attribute of interest, or several types of attributes associated with each feature. When several types of attributes are associated with the spatial features, the data are called multivariate. It is vital to distinguish spatial (location space) from multivariate (variable space), since spatial data can be referenced by two coordinates in the plane that may also be considered "variables. Multivariate refers to more than one type of attribute; multidimensional refers to more than one coordinate axis in space. Thus, spatial data consist of features indexed by spatial locations and with specified supports and attributes associated with those features. Spatial statistical analysis is based not only on the attribute data, but also depending on the spatial locations and the features associated with these locations.

3.2 Types of spatial data

3.2.1 Points

A point pattern dataset gives the locations of objects/events occurring in a study region. The points could represent trees or landmarks. Points might be situated in a region of the two-dimensional (2D) plane, or on the Earth’s surface, or a 3D volume. They could be points in space-time (e.g. earthquake epicenter location and time).

3.2.2. Marks

The points may have extra information called marks attached to them. The mark represents an “attribute” of the point. The mark variable could be categorical, e.g. species or disease status: or continuous, e.g. individual heights or weights.
3.2.3 Covariates

Spatial data may also include covariates—any data that we treat as explanatory, rather than as part of the 'response'. Covariate data may be a spatial function $Z(u)$ defined at all spatial locations $u$, e.g. altitude, disease status, displayed as a pixel image or a contour plot. Covariate data may be another spatial pattern such as another point pattern, or a line segment pattern, e.g. a map of disease intensity.

3.3 Exploratory Data Analysis

3.3.1 Tests of Complete Spatial Randomness

Under complete spatial randomness (CSR), points are independent of each other and have the same propensity to be found at any location. Testing for CSR is an important part of exploratory data analysis of point patterns: if the null hypothesis is accepted, one can assume that the given point pattern is completely spatially random, which has two main consequences:

(a) There is no need to consider a more complicated model; the simple Poisson process model can be used.

(b) If no additional information or data on underlying processes is available it is not possible to find indicators of interesting interaction between the points based on the geometry of the observed pattern alone.

Its basic properties are: the number of points falling in any region $A$ has a Poisson distribution with mean $\lambda \cdot \text{area}(A)$; given are $n$ points inside region $A$, the locations of these points are i.i.d. and uniformly distributed inside $A$ and the contents of two disjoint regions $A$ and $B$ are independent.
The uniform Poisson process is often the 'null model' in an analysis.

The **homogeneous Poisson process** of intensity \( \lambda > 0 \) has the properties

- **PP1**: the number \( N(X \cap B) \) of points falling in any region \( B \) is a Poisson random variable;
- **PP2**: the expected number of points falling in \( B \) is \( \mathbb{E}[N(X \cap B)] = \lambda \cdot \text{area}(B) \);
- **PP3**: if \( B_1, B_2 \) are disjoint sets then \( N(X \cap B_1) \) and \( N(X \cap B_2) \) are independent random variables;
- **PP4**: since \( N(X \cap B) = n \), the \( n \) points are independent and uniformly distributed in \( B \).

\( \text{(PP2)} \) and \( \text{(PP3)} \) are sufficient.

The **inhomogeneous Poisson process** with intensity function \( \lambda(u), u \in \mathbb{R}^2 \), is a modification of the homogeneous Poisson process, in which properties (PP2) and (PP4) above are replaced by:

- **(PP2')**: the number \( N(X \cap B) \) of points falling in a region \( B \) has expectation

  \[ \mathbb{E}[N(X \cap B)] = \int_B \lambda(u) \, du \]

- **(PP4')**: given that \( N(X \cap B) = n \), the \( n \) points are independent and identically distributed, with common probability density \( f(u) = \lambda(u)/I \), where \( I = \int \lambda(\mu) \, d\mu \).

The intensity argument \( \lambda \) can be a constant, a function \((x,y)\) giving the values of the intensity function at coordinates \((x,y)\), or a pixel image containing the intensity values at a grid of locations.
The **log-likelihood** for the **homogeneous Poisson** process with intensity $\lambda$ is

$$\log L(\lambda; x) = n(x) \log \lambda - \lambda \text{Area}(W),$$

where $n(x)$ is the number of points in the dataset $x$ and $W$ is the spatial weight denoting strength of interaction.

The maximum likelihood estimator of $\lambda$ is

$$\hat{\lambda} = \frac{n(x)}{\text{Area}(W)},$$

which is also an unbiased estimator.

Where $W$ is the spatial weight set denoting the strength of the connection between areas

The variance of $\hat{\lambda}$ is

$$\text{var}(\hat{\lambda}) = \frac{\hat{\lambda}}{\text{Area}(W)}.$$

Consider an inhomogeneous Poisson process with intensity function $\lambda_\theta(\mu)$ depending on a parameter $\theta$.

The log-likelihood for $\lambda$ is

$$\log L(\lambda; x) = \sum_{i=1}^n \log \lambda_\theta(x_i) - \int \lambda_\theta(\mu) d\mu.$$

If $\log \lambda_\theta(\mu)$ is linear in $\theta$, then the log-likelihood is concave, so there is a unique maximum likelihood estimator (MLE). However, the MLE $\hat{\theta}$ is not analytically tractable, so it must be computed using numerical algorithms such as Newton’s method. The usual asymptotic theory of maximum likelihood applies: under suitable large sample conditions, the MLE of $\theta$ is asymptotically normal. If we wish to test CSR, the likelihood ratio test statistic

$$R = 2 \log \frac{L(\hat{\theta})}{L(\hat{\lambda})},$$

is
is asymptotically $x^2$ under CSR, and this gives an asymptotically optimal test of CSR against the alternative of an inhomogeneous Poisson process with intensity $\lambda_q(\mu)$.

3.3.2 The G-Function: Nearest Neighbor Distance Distribution

Achia (2009) while quoting (Cressie, 1991; Diggle, 1983; Ripley, 1988) defines the G-function as the nearest neighbor distance distribution function (also called the event-to-event" or inter-event" distribution) of a point process $X$ and is the cumulative distribution function $G$ of the distance from a typical random point of $X$ to the nearest other point of $X$. An estimate of $G$ derived from a spatial point pattern dataset can be used in exploratory data analysis and formal inference about the pattern. In exploratory analyses, the estimate of $G$ is a useful statistic for summarizing one aspect of the clustering of points.

If these distances are defined as $d_i = \min_j \{d_{ij}, \forall j \neq i\}, i = 1, \ldots, n$, then the $G$ function can be estimated as

$$\hat{G}(r) = \frac{\{d_i; d_i \leq r, \forall i\}}{n},$$

where the numerator is the number of elements in the set of distances that are lower than or equal to $d$ and $n$ is the total number of points. Under CSR and for inferential purposes, the estimate of $G$ is usually compared to the true value of $G$ for a completely random (Poisson) point process, which is

$$G(r) = 1 - \exp(\lambda \pi r^2)$$

where $\lambda$ is the intensity (expected number of points per unit area).

The compatibility with CSR of the point pattern can be assessed by plotting the empirical function $\hat{G}(d)$ against the theoretical expectation. In addition, point-wise envelopes under CSR can be computed by repeatedly simulating a CSR point process with the same estimated
intensity \( \hat{\lambda} \) in the study region (Diggle, 2003) and checking whether the empirical function is contained within the study area. Deviations between the empirical and theoretical G curves may suggest spatial clustering or spatial regularity.

Assuming the point process X is stationary, the cumulative distribution function of the nearest-neighbor distance for a typical point in the pattern can be defined as,

\[
G(r) = P\{d(u, X \backslash \{u\}) < r \mid u \in X\},
\]

where u is an arbitrary location, and \( d(u, X \backslash \{u\}) \) is the shortest distance from u to the point pattern X excluding u itself. If the process is stationary then this definition does not depend on u. The empirical distribution function of the observed nearest-neighbor distances

\[
G^*(r) = \frac{1}{n(x)} \sum_i \{t_i \leq r\},
\]

is a negatively biased estimator of G(r).

Many edge corrections are available. Typically they are weighted versions of the expected cdf,

\[
\hat{G}(r) = \sum e(x_i, r) \{t_i \leq r\},
\]

where \( e(x_i, r) \) is an edge correction weight designed so that \( \hat{G}(r) \) is approximately unbiased. To test for the G-function using ArcGIS 9.3, the geographically weighted regression which tests for inverse weighted distance was used.

### 3.3.3 Geographically Weighted Regression (Inversely Weighted Distance)

Geographically Weighted Regression is a technique for exploratory spatial data analysis. In "normal" regression it is assumed that the relationship we are modeling holds everywhere in the study area i.e. the regression parameters are "whole-map" statistics. In many situations this is not necessarily the case, as mapping the residuals (the difference between the observed and
predicted data) may reveal otherwise (Anselin, 1992). Regression-based models largely ignore this assumption, much to the detriment of spatially varying relationships.

A 'normal' regression model with one predictor variable can be written:

\[ y = \beta_0 + \beta_1 x + \epsilon \]

where

\( y \) = dependent variable,
\( x \) = independent variable,
\( \beta_0 \) & \( \beta_1 \) = parameters to be estimated, and
\( \epsilon \) = random error term (assumed to be normally distributed).

The assumption is that the values of \( \beta_0 \) & \( \beta_1 \) are constant across the study area. This means that if there is any geographic variation in the relationship then it is confined to the error term. GWR helps to treat this relationship as a non residual (Mitchell, 2008)

However, using data points with coordinates the above model can be re-written as:

\[ y(u,v) = \beta_0(u,v) + \beta_1(u,v)x_1 + \epsilon(u,v) \]

In this study, GWR is organized such that data points nearer HIV positive outcomes are given a heavier weight in the model than data further away. The resulting parameter estimates are then mapped as intensity functions.
3.3.4 Identifying Outliers

Geo-statistical methods are sensitive to outlying values that exert a significant effect on predictions. Extreme outliers were therefore identified and excluded using a spatial filter. The method assumes that the probability of an unusually large variance in geo-coordinate value being a genuine 'outlier' is larger if (a) it is in a neighborhood of generally much smaller values and/or (b) the neighborhood is generally uniform.

Statistical outliers were identified using a spatial filtering algorithm that implemented the following procedure for each datum \( p(x_i) \) at location \( x_i \).

1. Around each data location, \( x_i \), a regional neighborhood was specified defined as the set of \( n_i \) data \( \{p(x_j); j = 1,...,n_i\} \) located within five decimal degrees of \( x_i \). This set excluded the datum \( p(x_i) \) itself.

2. For each of the \( (n_i^2 - n_i)/2 \) unique pairs of data \( \{(p(x_j), p(x_{j'}))\} \) within the regional neighborhood, the semi-variance \( \gamma_{j,j'} = 0.5(p(x_j) - p(x_{j'}))^2 \) was computed along with their separation distance \( h_{j,j'} \). This set of \( (n_i^2 - n_i)/2 \) pair-wise values of \( \gamma_{j,j'} \) and \( h_{j,j'} \) calculated without reference to the datum \( p(x_i) \) represented the underlying spatial autocorrelation structure within the regional neighborhood around location \( x_i \).

3. The next step was to compare the spatial statistical similarity between the datum \( p(x_i) \) and it neighbors with the corresponding underlying relationships represented by the regional semi-variance values. Semi-variances \( \gamma_{i,k} \) were calculated between the datum \( p(x_i) \) and its 20 nearest neighbors \( \{p(x_k); k = 1,...,20\} \). Each of these 20 semi-variance values \( \gamma_{i,k} \) were then compared to regional semi-variance values \( \gamma_{i,j'} \) computed for pairs.
at similar separation distances, defined as within 0.1 decimal degrees. This required the definition of the subset \( \gamma_{j,j'}^k = \{ \gamma_{j,j'} : |h_{j,j'} - h_{i,k}| < 0.1dd \} \) for each of the semi-variance values \( \gamma_{i,k} \).

4. The quantile position \( q_i \) of each semi-variance value \( \gamma_{i,k} \) within the corresponding subset \( \gamma_{j,j'}^k \) was recorded and the mean of the 20 quantile positions \( q_i = \frac{1}{20} \sum q_k \) was calculated.

The value \( q_i \) provides, for every datum \( P(X_i) \), an index of the extent to which the observed survey datum value differs from its neighbors relative to the extent expected given the distances to those neighbors, as defined by the regional semi-variances. Data for which the value \( q_i \) exceeded 0.95 were identified as outliers.

### 3.4 Variable description

**HIV Test Result:** This is the test outcome that is either positive or negative. It is also the dependent variable

**Sex:** The sex of the respondents (male or female)

**Age:** The age of the respondents. This was further categorized into 4 age groups

**Marital status:** This is marriage status of the study respondents. There are three wide categories of unmarried, married and individuals who were married but aren’t married anymore. The unmarried category is comprised of single, steady partner not living together and steady partner living together. The married group is comprised of those who are in
monogamous and polygamous unions while those who are no longer married group is made of divorced or widowed persons.

**Highest level of Education:** The highest level of education achieved by the respondent. There are four categories: None for those who never attended any formal education; primary for those who had basic education; secondary for individuals who had post primary education and college/university group comprises individuals who have furthered their education beyond secondary level.

**Occupation:** This variable describes what socio-economic activity individuals are involved in. Students are defined as those who do not involve themselves in any other fulltime activity other than academic work. Employed individuals were those who are either in any form of casual job engagement or full time salaried engagement. The self employed are those who operate their own income generating activities or businesses. The unemployed are individuals not engaged in any socio-economic activity.

**3.5 Spatial analysis**

Spatial analysis began as early as during the geo-locating of residential addresses of respondent homesteads. Data analysis was conducted using ArcGIS Version 9.3 software. Exploratory data analysis was done using homogeneous and inhomogeneous Poisson processes, nearest neighbor distance distribution function and Geographically Weighted Regression statistics. Autocorrelation was done using ordinary least squares to obtain the Poisson point estimates while spatial logistic regression for estimates of HIV outcome against the five demographic
features. Of 13,060 respondents, 412 (3%) locations were outside of the area of study, could not be located, or could not be unambiguously mapped.

The data was explored for significant spatial clustering of demographic characteristics and HIV outcome. Spatial autocorrelation, i.e. the correlation of HIV outcome between pairs of neighboring observations, was the core form of the exploratory data analysis (Tukey, 1977). 'Neighboring observation' is the creating of a contiguity or distance matrix in which the correlation between neighbors is compared with the general variance of the sample as in ordinary correlation analysis.

3.6 Models

3.6.1 Point level Models

For point-level data, the location index, $S$, varies continuously over $D$, a fixed subset of $\mathbb{R}^d$. If the covariance between the random variables at two locations is assumed to depend on the distance between the locations, then an exponential model of association would be used. Here the covariance between measurements at two locations is an exponential function of the inter-location distance, i.e.,

$$Cov[Y(s_i), Y(s_j)] = C(d_{ij}) = \delta^2 e^{-d_{ij}/\phi},$$

where $i \neq j$ is the distance between sites $S_i$ and $S_j$, and $\delta^2$ and $\phi$ are positive parameters called the partial sill and the decay parameter, respectively ($1/\phi$ is called the range parameter). A plot of the covariance versus distance is called the covariogram. When $i = i', \ d_{ii'}$ is 0, and $C(d_{ii'}) = \text{var}(Y(s_i))$ is often expanded to $\nu^2 + \delta^2$, where $\nu^2 > 0$ is called a nugget effect and $\nu^2 + \delta^2$ is called the sill.
3.6.2 Point Process Models

In the point process model, the spatial domain $D$ is itself random, such that the elements of the index set $D$ are the locations of random events that constitute the spatial point pattern. $Y(s)$ then equals the constant 1 for all $s \in D$ (indicating occurrence of the event), but it may also provide additional covariate information equivalent to a marked point process.

The most commonly used homogeneous Poisson process (CSR) measure its spatial independence is Ripley's $K$ Function.

The $K$ function is

$$K(t) = \lambda^{-1} E \left( \text{number of extra events within distance } t \text{ of a randomly chosen event} \right)$$

where $\lambda$ is the density (number per unit area) of events and $K(t)$ describes characteristics of the point processes at many distance scales. Alternative summaries (e.g. mean nearest-neighbor distance or the cumulative distribution function) of distance from random points to their nearest neighbors do not have this property.

3.6.3 Spatial Models using Akaike Information Criterion (AIC)

There are often model selection criterions: the first presumes that there exists a true finite-dimensional model from which the data were generated. The hypothesis is that the true model is linear and there exists an explicit linear relationship between the explanatory variables and the response. In this case, the key modeling objective is to identify the correct set of covariates that comprise the model. The alternative modeling criterion, assumes that the underlying true model is infinitely dimensional and it is not possible to identify all the requisite factors that go into the process under study (Hoeting et al, 2004).
Under the first scenario, consistency should be a minimum requirement of a model selection procedure. That is, as more data are acquired, the model selection procedure should ultimately choose the correct model with probability one. In the second situation when the true model is infinite dimensional, a model selection procedure ought to choose a finite dimensional model that is closest to the true model in some sense. The Akaike Information Criterion (Akaike, 1978) is one procedure that is designed to achieve this second goal. It is the measure of the loss of information incurred by fitting an incorrect model to the data. Model selection techniques for spatial models include the correlation structure in determining the best set of predictors. By computing the AIC of all possible sets of explanatory variables and autocorrelation functions, a single “best” model or a set of models which fit the data well can be obtained (Haining, 1990). A model with the lowest AIC is therefore considered the best.
Chapter Four: Results and Discussion

4.1 Descriptive and Multivariate Logistic Regression Analysis of the risk factors for HIV

Without adjusting for the 3% outliers, 13,060 persons were counseled and/or tested over the six month HBHTC campaign, the ratio of males to females was 1:1 (6532:6528). HIV prevalence was approximately twice in females as was observed in males (3.0% vs 1.6%). Overall, 11,582 (88.7%) persons were in the sexual age group of 15-54 years. The mean (median) age was 27.5 (25, IQR 22, 34) years for males and 32.3 (26.1, IQR; 21, 30) for females. The mean (median) age was 27.5 (25, IQR 22, 34) years for males and 32.3 (26.1, IQR; 21, 30) for females.

A total of 6,254 (47.9%) participants were either in monogamous or polygamous marriages, 2,654 (20.3%) were in a relationship while 812 (6.2%) were either divorced or widowed. Even though divorced and widowed individuals formed a cumulative 6.2% of the total sample, they constituted 26% of total individuals infected by HIV. 66% (194) of all HIV infected persons were either self employed or unemployed. Individuals with basic (primary) education constituted the bulk of persons with HIV (65.1%) while those with post secondary education were least afflicted by HIV (6.25% of all cases).

The analysis of the association of the risk of HIV and other potential risk factors is presented in Table1 below.

From the logistic model, it estimates that there was a high statistically significant association between marital status and HIV, p=0.001 based on a multivariate Wald statistic Chi-Square(6) =30.13 Since in this population the prevalence of HIV is approximately 2.3%,
the estimated odds ratios were interpreted as risks. Overall, the observed respective odds (risks) of HIV for divorced, widowed and married polygamous are 3.947, \( p=0.0000 \) and (95% CI: 2.418-6.45); 3.368, \( p=0.0000 \) and (95% CI: 2.04-5.56); and 2.041, \( p=0.0002 \) and (95% CI: 1.40-2.983), times higher than the odds (risk) of HIV for individuals who were in steady relationships but not married. Primary education and unemployment were generally associated with a statistically significant 62% and 41% increase in the risk of HIV (\( OR=0.376, \ p=0.0013, \ 95\% \ CI: 0.2077-0.6817 \) and \( OR= 0.585, \ p=0.0004, \ 95\% \ CI: 0.436-0.786 \) respectively.

Table 1: Descriptive and Multivariate Logistic Regression Analysis of the risk factors for HIV

<table>
<thead>
<tr>
<th>Variable</th>
<th>HIV+</th>
<th>HIV-</th>
<th>OR</th>
<th>95% Confidence Interval for Exp(B)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>100</td>
<td>6432</td>
<td>1.8900</td>
<td>1.1200</td>
<td>4.3479</td>
</tr>
<tr>
<td>Females</td>
<td>192</td>
<td>6336</td>
<td>1</td>
<td>1.2900</td>
<td>4.5479</td>
</tr>
</tbody>
</table>

| Education         |      |      |      |                                   |      |
| None              | 10   | 302  | 1.3763 | 0.2077 | 0.6817 | 0.0013 |
| Primary           | 190  | 5916 | 0.5328 | 0.2219 | 1.2795 | 0.1590 |
| Secondary         | 78   | 5314 | 0.7254 | 0.3924 | 1.3409 | 0.3058 |
| Post-Secondary    | 12   | 1158 | 0.7254 | 0.3924 | 1.3409 | 0.3058 |

| Marital Status    |      |      |      |                                   |      |
| Single            | 36   | 3296 | 1     |                                   |      |
| Steady-not living together | 30   | 2160 | 1.1408 | 0.6208 | 2.0966 | 0.6713 |
| Steady-living together | 10   | 454  | 0.6397 | 0.3320 | 1.2327 | 0.1819 |
| Married monogamous | 150  | 5756 | 2.4656 | 1.2025 | 5.0557 | 0.0138 |
| Married polygamous | 16   | 332  | 2.0413 | 1.3964 | 2.9838 | 0.0002 |
| Divorce           | 14   | 146  | 3.9473 | 2.4175 | 6.4451 | 0.0000 |
| Widowed           | 36   | 616  | 3.3681 | 2.0396 | 5.5620 | 0.0000 |

| Occupation        |      |      |      |                                   |      |
| Student           | 8    | 1244 | 1     |                                   |      |
| Employed          | 86   | 3440 | 1.4733 | 0.6812 | 3.1863 | 0.3248 |
| Self-employed     | 106  | 3754 | 0.9941 | 0.7266 | 1.3600 | 0.9705 |
| Unemployed        | 88   | 4168 | 0.5852 | 0.4360 | 0.7856 | 0.0004 |
4.1a Adjusted Binary Logistic Regression Analysis of the risk factors for HIV

Females were 1.75 times at risk than males (p=0.000, 95% CI, 1.36-2.26). From the adjusted binary logistic model, those who did not have any education were 1.83 odds to contract HIV compared to 0.62 odds for individuals who had post-primary education respectively (p=0.000, 95% CI, 1.41-2.38 and p=0.161, 95% CI, 0.322-1.31) respectively. Divorcees and widowers were 3.16 odds to contract HIV against married individuals who were 1.48 odds (p=0.000, 95% CI, 2.09-4.76 and p=0.011, 95% CI, 1.10-2.00) respectively. Employment was associated with a 1.5 times more risk of contracting HIV. (p=0.006, 95% CI, 1.11-1.83).

Table 2: Adjusted Binary Logistic Regression Analysis of the risk factors for HIV

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Sig.</th>
<th>OR</th>
<th>95.0% C.I. Lower</th>
<th>95.0% C.I. Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>-0.559</td>
<td>0.131</td>
<td>18.353</td>
<td>0.000</td>
<td>1.748252</td>
<td>1.355014</td>
<td>2.257336</td>
</tr>
<tr>
<td>Age</td>
<td>0.022</td>
<td>0.005</td>
<td>16.214</td>
<td>0.000</td>
<td>0.978474</td>
<td>0.968054</td>
<td>0.98912</td>
</tr>
<tr>
<td>No Education</td>
<td>-0.606</td>
<td>0.132</td>
<td>21.044</td>
<td>0.000</td>
<td>1.831502</td>
<td>1.414427</td>
<td>2.375297</td>
</tr>
<tr>
<td>Post-Primary</td>
<td>0.472</td>
<td>0.336</td>
<td>1.965</td>
<td>0.161</td>
<td>0.62383</td>
<td>0.322685</td>
<td>0.926217</td>
</tr>
<tr>
<td>Married</td>
<td>-0.391</td>
<td>0.153</td>
<td>6.51</td>
<td>0.011</td>
<td>1.47929</td>
<td>1.09529</td>
<td>1.996008</td>
</tr>
<tr>
<td>Divorced</td>
<td>-1.151</td>
<td>0.21</td>
<td>29.971</td>
<td>0.000</td>
<td>3.164557</td>
<td>2.09205</td>
<td>4.761905</td>
</tr>
<tr>
<td>Employed</td>
<td>-0.355</td>
<td>0.129</td>
<td>7.644</td>
<td>0.006</td>
<td>1.426534</td>
<td>1.108647</td>
<td>1.834862</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.024</td>
<td>480</td>
<td>39.664</td>
<td>0.000</td>
<td>20.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

46
4.2 Classification Decision Rule

Table 3: Classification Decision Rule

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HIV</td>
<td>12530</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>286</td>
<td>0</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The specificity and sensitivity of the predictors are 100% respectively. The predictors also give 0% false positive and 2.23% false negative. Hosmer Lemeshow goodness of fit test with a stepwise method at 0.25 entry and 0.03 removal at 95% CI show an overall 97.8% prediction.

4.3 Point Pattern of HIV Prevalence

Figure 4: Point Patterns of HIV Prevalence
Figure 4 above shows the spatial distribution of all the respondents against their HIV outcome. Positive HIV outcomes show slight clustering. It is a spatial representation of HIV outcomes within the larger Kalimoni sub-location. Points outside Thika district have been considered outliers. Graph 8 below shows the distribution of HIV outcomes across the study location.

4.4 Geographically Weighted Regression (Inverse Distance Weighting)

*Figure 5: Geographically Weighted Regression (Inverse Distance Weighting)*

This graph shows four clusters in which there are high concentrations of HIV positive outcomes. These are Ruiru, Kiandutu, Juja Farm and Munyumunyu. Ruiru is cosmopolitan urban centre with a high population density mainly comprised of students (universities, colleges and
secondary schools) local inhabitants, employees of industries in and around Ruiru town. The high HIV positive outcomes can be partially attributed to the high population

Kinandutu is a slum informal settlement characterized by high population density. Majority of the inhabitants here lie in temporary shelter, have meager sources of income and therefore unable to meet most of their basic needs. A combination of these circumstances could partly explain the high concentration of HIV positive outcomes in this area which could a consequence of the inhabitants engaging in risky sexual behaviors. High HIV prevalence rates have been reported in other slum areas in Nairobi such as Viwandani and Korogogocho (Madise, et al 2003).

On the other hand, Juja Farm is mainly comprised of casual laborers working in a local commercialized agricultural industry. Of the 160 married respondents staying around Juja Farm, only 9% lived with their spouses. The laborers are likely therefore to have engaged in risky sexual behaviors hence the high HIV positive outcomes.

Ten neighbors effect were used for each HIV positive outcome’s locality to establish neighborhood effects and control for the degree of smoothing in the GWR. Using a Kernel type bandwidth, the corrected Akaike Information Criterion (AIC) used for identifying optimal fixed distance or adaptive neighbors HIV outcome was 970 meters (approximately 1 km). The function of the number of nearest neighbors such that each HIV positive outcome’s estimation was based on this bandwidth instead of a specific distance. The threshold distance for HIV positive estimation was 0.135 degrees (approximately 15 kms).
Results indicate that females were twice likely to contract HIV compared to their male counterparts (0.029:0.015). OLS results show a high significant relationship between sex and HIV outcome ($p=0.0000$). Odds ratio output computed using SPSS indicate that females were two times more at risk of contracting HIV than males ($OR=1.89$, 95% CI, 1.12-3.348). These results are consistent with findings in a number of past studies undertaken in Kenya. The KDHS (2003) and KAIS (2007) revealed that 3 out of every 5 HIV infected individuals are women. In their study to determine why young women have a much higher prevalence of HIV than young
divorced exhibited the highest probability of contracting HIV at 3.6%, 4.3% and 5% respectively. 3 out of 100 individuals in monogamous marriages were likely to be HIV positive.

*Figure 7: Ripley’s K Function*

4.5.3 *Figure 8: HIV outcome and Age groups*
and 2.2% respectively of being HIV positive. Respondents in polygamous unions, widowed or divorced exhibited the highest probability of contracting HIV at 3.6%, 4.3% and 5% respectively.

3 out of 100 individuals in monogamous marriages were likely to be HIV positive. In their study to investigate the relationship between marital status and risk of HIV in males and females aged above 15 years in South Africa, Shisana et al (2004) say that available evidence on the relationship between marital status and HIV is contradictory. They conclude that the risk depends on various demographic factors and sex behavior practices. In their study to determine risk factors for recent HIV infection in Uganda, Mermin et al (2008) reported that current marital status and positive HIV status (widowed vs never married, OR, 6.1; 95% CI, 2.8-13.3; divorced vs never married, OR, 3.0; 95% CI, 1.5-6.1) do not clearly show which marital status can be clearly associated with high risk of HIV infection without looking at other factors that directly or indirectly influence or cause HIV.
4.5.3 HIV outcome and Age

Figure 8: HIV outcome and Age groups

Figure 11 above shows HIV prevalence by age groups. Results show children below 15 years and individuals above 55 years were the least likely to be HIV positive. The probability of 15-24, 25-34, 35-55 years' age group show a steady rise with probabilities 0.026, 0.038 and 0.05 respectively. This trend is consistent with the KDHS 2003 and KAIS 2007 reports. The peak HIV prevalence in Kenya for females is between ages 30-34 while that for men is between ages 40-44 years (KAIS).
4.5.4 HIV outcome and Highest Level of Education

Results in Figure 12 above, show lowest likelihood of HIV infection among individuals with tertiary education (probability=0.006) while the highest likelihood is found in individuals who had basic (primary) education (probability =0.041). However, OLS analysis does not show any significant spatial relationship between level of education and HIV outcome (p=0.057). Respondents who had secondary level education were approximately twice (probability =0.029) more likely to contract HIV than those who had not had any formal education (probability =0.017). In this study, the low probability of respondents who did not have formal education could have been partly affected by presence of children of non-schooling age who were also
tested with the consent of their parents or guardians. Most studies from sub-Saharan Africa suggest a positive association between educational attainment and HIV infection, i.e. more educated at higher risk. However, some studies conducted in matured epidemics have reported a shift towards reduced risk differentials (Michelo & Filkesnes, 2004).

4.5.5 HIV outcome and Occupation

Figure 10: HIV outcome and Occupation

Results from Figure 13 show that casual laborers were at the highest risk of contracting HIV (probability=0.024). Students and salaried employees were least likely to be HIV positive. However the OLS test does not establish any significant relationship between occupation and HIV test outcome among the respondents in the study (p=0.0845).
4.6 Model selection

After adjusting for outliers and running the binary logistic regression, the below equation was the best predictors of risk of contracting HIV in Kalimoni sub-location:

\[ Y = -3.03 - 1.51x_1 - 0.61x_2 - 0.56x_3 - 0.36x_4 + 0.022x_5 \]

where \( x_1 \) is divorced or widowed marital status, \( x_2 \) is no education at all, \( x_3 \) is female respondents, \( x_4 \) is employment and \( x_5 \) is age.

4.7 Discussion

Results from Kenya AIDS Indicator Survey (KAIS) indicate that 7.4 % of Kenyan adults age 15-64 are infected with HIV, the virus that causes AIDS. This shows that more than 1.4 million Kenyans are living with HIV/AIDS. In 2003, KDHS estimated a prevalence of 6.7 % among 15-49 year olds. For the same age group, KAIS estimates that 7.8 % are infected. KAIS revealed higher proportion of women age 15-64 (8.7 %) than men (5.6 %) are infected with HIV. This pattern is similar to what was observed in 2003. This means that 3 out of 5 HIV-infected Kenyans are female (KAIS 2007).

Women face considerably higher risk of HIV infection than men, and also experience a shorter life expectancy due to HIV/AIDS. The 7th edition of AIDS in Kenya reports an HIV prevalence rate of 8% in adult women and 4% in adult men. Populations in Kenya especially at risk include injecting drug users and people in prostitution, whose prevalence rates are estimated at 53% and 27%, respectively (Wikipedia Kenya Profile, 2008)

Increase in age within the sexually active individuals results in increased risk of HIV infection.
4.8 Ethics

1. Consent was obtained prior to counseling and testing, including consent to follow up positive clients who did not attend referral points.

2. For the purpose of protecting participant confidentiality, a small random error has been incorporated into the geographical position of each participant. This also serves to create an accurate visual representation of the distribution of all participants in the survey due to multiple participants sometimes being resident in a single homestead.
The study identified spatial clusters within Kalimoni sub-location after controlling for spatial auto-correlation and other important risk factors. Although sharing the same pattern of education and occupation effects on the risk of HIV infection, the greatest disparity in HIV prevalence is experienced by sex, age and marital status. The study demonstrates the power of the spatial approach, in teasing out clustering in even hitherto homogeneous settings.

Lack of resources continues to hamper the development and use of GIS in health care interventions. However, individual organizations have made specific efforts to use GIS to improve on their service delivery and quality of life programs. Recent achievements in use of GIS in healthcare interventions are the regular trainings organized for service providers, increased mapping service activities, increased demand for geospatial data for geographic understanding of the pattern of the HIV epidemic and targeted effective interventions.

Increased prevention strategies that take spatial contexts into account are needed for efficient messaging and effective service delivery. Despite the overall low prevalence of HIV in Kalimoni sub-location, the results support the need for interventions that target socio-geographic spaces (clustered villages) at greatest risk to supplement measures aimed at the general population.

The use of spatial surveillance of HIV infection and prevalence approaches is key to monitoring the changing epidemiology of HIV in sub-populations. The use of geo-statistical methods can
help focus surveillance efforts and define areas where uncertainty exists, guiding area specific interventions because of better estimates of where people live, disease burden and how these might change with changing infection-risk exposure.

Further Research
The analysis presented in this report is a cross tabulation of HIV outcome/risk and demographic factors. Further research exploring unique combinations of spatial, demographic, socio-economic, behavioral, biological & co-infection and physical (e.g. roads or healthcare facilities) factors could yield better and precise results.

Study Limitations
412 (3%) locations were either outside the area of study, could not be located, or could not be unambiguously mapped. The data collectors were not experts in GIS or survey to accurately map study locations and boundaries as indicated in the Central Bureau of Statistics (CBS) maps.

Boundaries of the study area have changed considerably. Whereas the District AIDS and STI's Coordinator (DASCO) have their defined areas of jurisdiction, these areas do not coincide with the current provincial administrative areas of jurisdiction. Consequently there is overlap between Ministry of Health boundaries and provincial administration boundaries.

The selection of participants in this study was not a random process. Clients were 'recruited' on a consenting basis. This could have led to sampling bias.
References


APPENDIX

Appendix 1: Script Ordinary Least Squares

Figure 11: Ordinary Least Squares

# Analyze existence of spatial autocorrelation in residuals of HIV +ve points
# Data from CBHCT data using Ordinary Least Squares Regression

# Import system modules
import arcgisscripting

# Creating the Geoprocessor object
gp = arcgisscripting.create(9.3)
gp.OverwriteOutput=1

# Importing data
workspace="C:\\MSc Project\\Data\\Final Data"
gp.workspace=workspace

# HIV Result as a function of (log of sex, marital status, education, Occupation)
# Process: Ordinary Least Squares
ols=gp.OrdinaryLeastSquares("FinalData.shp", "HIV Result",
"olsResults.shp", "HIVResults",
"LOGHIVR;FinalData; sex",
"olsCoefTab.dbf",
"olsDiagTab.dbf")

"olsResults.shp", "HIVResults",
"LOGHIVR;FinalData; maritalstatus",
"olsCoefTab.dbf",
"olsDiagTab.dbf")

"olsResults.shp", "HIVResults",
"LOGHIVR;FinalData; education",
"olsCoefTab.dbf",
"olsDiagTab.dbf")

"olsResults.shp", "HIVResults",
"LOGHIVR;FinalData; occupation",
"olsCoefTab.dbf",
"olsDiagTab.dbf")

# Create Spatial Weights Matrix
# Process: Generate Spatial Weights Matrix
swm=gp.GenerateSpatialWeightsMatrix("Results.shp", "HIVResults",
"euclidean6Neighs.swm",
"G_NEAREST_NEIGHBORS","#", ", #", 10)

# Calculate OLS Index of Spatial Autocorrelation for
# OLS Residuals using a SWM File.
# Process: Spatial Autocorrelation
OLS=gp.SpatialAutocorrelation("olsResults.shp", "Residual",
"false","Get Spatial Weights From File",
"Euclidean Distance","None","#",
"euclideann10Neighs.swm")