Effect of sexual behaviour on the spatial distribution of HIV among Kenyan youth in 2003 and 2008

By

Brenda Anyango Oulo

A Research Report Submitted to the Faculty of Health Sciences, the University of the Witwatersrand in partial fulfilment of the requirements for the Degree of Masters in Epidemiology in the field of Infectious Disease Epidemiology.

March,2018.

Dedication

I would like to dedicate this research report to my family particularly my mother, Margaret Namirembe, to whom I am eternally grateful for her love, support and prayers. Asante sana for unconditionally supporting me through every stage of my academic journey. Your words of encouragement and constant belief in me have been a pillar of strength in difficult times.

And to my wonderful support system of friends especially Tsetse, Wendy and Daniele, you guys totally rock!

Declaration

I, Brenda Anyango Oulo, hereby declare that except where specific reference is made to the work of others, the contents of this research report titled "**Effect of sexual behaviour on the spatial distribution of HIV among Kenyan youth in 2003 and 2008**" are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This research report is the result of my work and includes nothing which is the outcome of work done in collaboration.

06/03/2018

Acknowledgements

And I would like to acknowledge the support and guidance of my supervisors, Prof Eustasius Musenge, Department of Epidemiology and Biostatistics, School of Public Health, University of the Witwatersrand and Mrs Lucy Chimoyi-Otwombe, Aurum Institute, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa.

Abstract

Background:

Kenya is among the fifteen countries contributing more than 75% of the global HIV prevalence and incidence. In these countries, clustering of risk behaviour across geographical space and population sub-groups such as young people contributes significantly to the disease burden. Low levels of knowledge on HIV and rejecting attitudes towards HIV infected people influence sexual risk behaviour among young people.

Objectives:

The primary aim of this research study was to determine and compare the relationship between sexual behaviour and the spatial distribution of HIV among young Kenyans across two-time periods, 2003 and 2008. The secondary aim was to investigate the direct and indirect determinants of HIV distribution.

Methods:

Secondary data analysis was performed using cross-sectional data obtained from two two-stage population based surveys conducted in Kenya. The main outcome investigated was HIV status among young people aged 15-24 years. The study employed Bayesian-based multivariable logistic regression models adjusting for both non-spatial and spatial random effects. A conceptual framework based on knowledge, attitudes and practices (KAP) was the basis of a generalised structural equation model (GSEM) fitted to investigate the complex relationship between sexual risk behaviour and HIV status.

Results:

Among 2,650 and 2,857 young people aged 15-24 years in 2003 and 2008 respectively, 3.6% and 2.9% were HIV positive with the majority being female. The level of comprehensive HIV knowledge increased from 42% to 78% between 2003 and 2008. In both years, there was county spatial variability in HIV prevalence among young Kenyan people. Having a partner with concurrent sexual partners had the greatest direct effect on the odds of HIV infection (adjusted OR 2.26; 95%CI 1.63-3.12), but this was indirectly affected by knowledge levels during the latter survey year.

Conclusion:

This study revealed significant spatial variability in the HIV prevalence among young Kenyan youth and significant association between HIV knowledge, high-risk sexual behaviour and HIV infection. Interventions targeting sexual behaviour in this age group should focus on the counties with the greatest disease burden before implementing a wider national approach.

Keywords: HIV, young people, spatial analysis, GSEM, Kenya.

Contents

List of Figures

List of Tables

Abbreviations

HIV	Human immunodeficiency virus
UNAIDS	Joint United Nations Programme on HIV/AIDS
SSA	Sub-Sahara Africa
PLWH	People living with HIV
OR	Odds ratio
DNA	Deoxyribonucleic acid
PCR	Polymerase chain reaction
AOR	Adjusted odds ratio
POR	Posterior odds ratio
CI	Confidence interval/ credible interval
DHS	Demographic and health survey
KDHS	Kenya demographic and health survey
GSEM	Generalized structural equation model
KAP	Knowledge, Attitudes and Practices
GIS	Geographic Information System
GPS	Global Positioning System
ESRI	Environmental Systems Research Institute

Chapter 1 Introduction

1.1 Background

The Human immunodeficiency virus (HIV) pandemic still poses a serious challenge to public health. Globally, there has been a 38% reduction in new HIV infections since 2001 although in recent years progress has slowed (1,2). Moreover, global averages may mask variations between and within countries (3). A recent report reveals that Kenya is among fifteen countries contributing greater than 75% of the global HIV prevalence (2). Even within Kenya, the geographical distribution of HIV infection is significantly skewed (4). In 2014, the majority (65%) of new HIV infections occurred in just nine of the forty-seven counties within the country (4).

Kenya's HIV prevention roadmap envisions a total reduction in new HIV infections by the year 2030 (4). The plan aims to achieve this through providing county governments with evidence-based strategic information for geographical and population-based prioritisation of HIV prevention efforts (4). Young people are among other priority populations such as discordant couples and sex workers targeted in the new approach because they are contributing up to 54% of new infections in the country (4).

The United Nations (UN) Secretariat defines young people or youth as those in the 15-24 years age group (5). Globally, they are also considered a high-risk population because they accounted for 34% of all incident adult HIV infections in 2015 (1,2).

Their heightened risk of HIV acquisition is due to their sexual risk behaviour among other factors (6).

In 2003 and 2008/9, Kenya conducted population-based nationally representative demographic and health surveys that assessed HIV knowledge, attitudes and sexual behaviour in addition to biological testing for HIV (7,8). The most recent national survey that included HIV testing was conducted in 2012 (9) however, this data has not been made available. This research study, therefore, utilises the available 2003 and 2008 data to investigate the effect of sexual behaviour among young people on HIV prevalence and the extent to which spatial variability can be attributed to spatial differences in risky sexual behaviour.

1.2 Literature review

1.2.1 Sexual risk behaviour among young people

Globally, an estimated 4 million young people are HIV infected, although, as already alluded to the burden varies substantially across continents, regions and even countries (1,2,10). Sub Saharan Africa (SSA) mainly Eastern and Southern Africa bears the biggest burden of the disease. Most countries in this region have generalised epidemics with infection well established among the general population (11). However, some population subgroups are disproportionately affected (1,2). Particularly worrying, is the high rates of incident infection among young people which are an impediment to global efforts to control the pandemic (1,2).

Worldwide, young people have been found to have strikingly similar beliefs and sexual practices in a systematic review of qualitative studies investigating factors that affect their sexual conduct (12). In the study, certain social factors and gender stereotypes were found to be comparable across many societies (12).

For example, men are expected to be highly heterosexually active and sexual experience is desired leading to multiple sexual partnerships. Conversely, women are supposed to be chaste prompting many young women to keep sexual encounters secret (12).

The gender stereotypes and peer pressure may lead to early sexual initiation among both young men and women. A global review of sexual behaviour identifies peer pressure as the single most important contributing factor to the first sexual encounter among a proportion of young men aged 16-24 years (13) Moreover, more than 40% of adolescent boys aged 15-19 years have had sexual intercourse according to a systematic review on the reproductive health of adolescents (14).

The 2011 UNICEF report indicates that on average 11% of young women have sexual debut before the age of 15 years although the proportions vary considerably among the world regions (15). The highest percentage was reported in Latin America (22%) while the lowest was in South Asia (8%) (15). In the latest nationally representative survey conducted in Kenya, young men were almost twice as likely to report sexual debut before age 15 years compared to young women (16).

In various research studies, the definition of early sexual debut varies considerably (17). In a systematic review of studies on early sexual debut as a risk factor for HIV acquisition in SSA, the definitions included age cut-offs ranging from 14 to 19 years (17). This research report adopts the UNAIDS definition of sexual debut before age 15 years (10).

The early initiation of sexual intercourse has been shown to increase HIV risk among women directly through biological effects and indirectly among both sexes through predisposition to subsequent risk-taking sexual behaviour (17).

In a cross-sectional study of the general Danish population, men who had early sexual debut were twice as likely to engage in subsequent high-risk behaviours such as having new multiple sexual partners within a six month period and intercourse with a commercial sex worker (18). Similarly, an analysis of two longitudinal studies in Seattle, United States of America (USA), revealed that early initiation of sexual activity and subsequent risky behaviour such as having sex while intoxicated and having many lifetime sexual partners heightens the risk of sexually transmitted infections among both men and women (19).

Besides a high number of lifetime sexual partners, partner concurrency has been demonstrated to have a more profound impact on HIV transmission at community level than practising serial monogamy (20,21). A simple definition of concurrency is overlapping sexual partnerships in which sex with one partner occurs between two acts of intercourse with a different partner, which may either be a long or short-term situation (22). According to a qualitative study among young heterosexual adults in the USA, concurrency is acceptable as long as partners establish expectations from the outset (23). Similarly, research done in Kenya suggests that among young men in ongoing sexual relationships, there is a 51% prevalence of concurrency and more than 30% prevalence among young women residing in urban areas (25).

In addition to the community level effects of concurrency on HIV transmission, further arguments pose that at the individual level HIV risk is significantly increased among the partners of people practising concurrency than among the individuals themselves (20,24). However, this may be highly subject to measurement errors from self-reporting of sexual behaviour as respondents may not readily divulge their engagement in concurrency but may be likely to report a partner's involvement.

Another sexual risk behaviour known to increase HIV transmission is transactional sex (11). A common definition refers to the exchange of gifts, money or other favours for sexual activity and distinction is made from commercial sex work (25). Young people particularly women are engaging in casual or non-casual transactional sexual relationships with older wealthier partners (26). Such encounters are unlikely to be protected using condoms increasing the risk of HIV transmission (11,26). Despite the fact that the bulk of available literature on transactional sex focuses on young women, some studies have sought to illustrate the involvement of young men (27,28). For example, young men from poor households in Malawi reported that in most instances, material or financial benefits were the sole reason for engaging in sexual relations with older women (29).

Along the shores of Lake Victoria in Kenya, transactional sex dubbed sex-for-fish among adults including young people is significantly contributing to the high HIV incidence and prevalence in the area (30–32). Additionally, poverty and declining access to fishing areas in the bays of the Lake have caused differential power dynamics in the transactional relationships.

Young women are compelled to yield to unprotected sex with fisherfolk to obtain fish for sale or food (33). Furthermore, another study conducted in the same area found that the majority of incident HIV cases occurred in young women under the age of 25 years who were engaging in transactional sex (34).

Empirical evidence attributes the differential HIV risk among young women and men to behavioural factors among other biological and sociocultural factors (35). The sexual relationships between young women and older men, ten or more years older, are among behavioural factors that have been proposed as an explanation for the disproportionate risk among young women. Young women cite financial and material gains as the primary incentive of engaging in cross-generational relationships with older men (36). Among this age group, there is a misconception that older married sexual partners present a lower risk of HIV than their peers who commonly engage in multiple partnerships (37).

The evidence linking cross-generational partnerships and HIV are inconclusive (37–41). Cross-sectional analyses of nationally representative surveys in South Africa and Kenya have found a higher risk of HIV acquisition among young women in age-discordant sexual relationships with older male partners (38,39). However, these findings are limited by the study design as longitudinal studies provide a more accurate estimation of risk. Contrastingly, two cohort studies in South Africa showed no association between a young woman having an age-disparate partner and HIV acquisition after controlling for sociodemographic factors such as educational attainment, marital and economic status as well as other behavioural factors such as condom use and casual or multiple sexual partnerships (40,41).

It is important to mention that information on the risk among young men in relationships with older women is scarce.

A study conducted in Kenya contradicts the suggestion that cross-generational partnerships are common in high prevalence areas (42). The study found a low prevalence of cross-generational relationships in Kisumu, an area with a high burden of disease along the shores of Lake Victoria (42). The researchers included marital status and economic asymmetry in the definition of cross-generational partnership. The term "sugar daddy" commonly used in Kenya to describe older male partners of younger women, was defined as a man 10 or more years older than his female partner in a non-marital partnership with a specified threshold of financial assistance (42).

The lack of consensus on cross-generational partnerships suggests the need for more research to conclude its effect on HIV acquisition decisively. Recent studies are further exploring the characteristics of age discordant relationships and the association with HIV (43,44). In one such study, the increased HIV risk observed in cross-generational partnerships was linked to sexual risk characteristics such as concurrency and inconsistent condom use (43).

1.2.2 Spatial variability of HIV distribution within countries

Spatial variations in HIV distribution within countries in SSA and elsewhere have been highlighted by several studies (3,45,46). A study in China revealed that the distribution of HIV cases varied significantly across the country with clustering in only four of the thirty-one provinces (44).

Furthermore, the available literature demonstrates that besides variations in prevalence at a particular point in time, evaluation across time periods may provide useful insights into the changing dynamics of the epidemic. In a spatial-temporal analysis of data from SSA, observed national prevalence declines were noted to obscure isolated increases or stagnation in HIV prevalence within clusters of high prevalence in Tanzania, Malawi, Kenya and Zimbabwe (3).

A similar study in Malawi which analysed antenatal clinic surveillance data from 1994 to 2010, not only revealed a wide variation in HIV spatial patterns at different administrative levels but also identified several trends within geographically defined epidemics (47). While the rural epidemic was relatively stagnant, the urban epidemic peaked early with a slow decline over the years compared to the semi-urban epidemic which varied considerably in the previous years (47).

Likewise, studies evaluating the spatial variability in HIV distribution in Kenya have found significant clustering of infection in the Western part of the country (30,45). An analysis using spatial effects to describe the HIV prevalence among the general population shows a high spatial heterogeneity of infection within a highly endemic area in Western Kenya (30). The researchers identified a significant HIV infected cluster along the shores of Lake Victoria using spatial scanning (30).

Moreover, a study evaluating the spatial distribution of sexual risk behaviour among young men in Kisumu, Kenya, found a significant cluster of young men who seldom used condoms in the six months before the study, the majority of whom were residing in a low-income area (48).

Other unique cultural practices such as widow inheritance and sexual cleansing of surviving partners after bereavement among some communities contribute to the high prevalence in the areas surrounding Lake Victoria (30). The high prevalence increases the likelihood of risky sexual encounters for young people in the region(30).

Apart from the fishing and surrounding communities, living in the capital city Nairobi, was positively and independently associated with greater odds of having recent HIV infection (49). After controlling for other factors such as demographics, sexual behaviour and biological factors, the likelihood of recent HIV infection was significantly increased by factors such as being a young uncircumcised male aged less than 30 years (adjusted odds ratio (AOR) 17.83; 95% CI 2.19–144.90) and having more than one sexual partner in the previous year (AOR 2.86; 95% CI 1.51–5.41) (49).

The characterisation of high HIV transmission areas, as well as sexual risk behaviour, is not only crucial for monitoring disease distribution but also to target specific interventions to where they are most needed. Targeting increases efficiency and effectiveness of intervention programmes (50).

1.2.3 Socio-demographic factors affecting HIV distribution

Consistently over several years, the evidence from countries across SSA reveal considerable age and gender disparities in HIV infection patterns among youth. In South Africa, young females aged 15-24 years have been shown to have the highest incidence of infection with the majority getting infected at least five to seven years earlier than their male counterparts (35,51,52).

A study among young people in rural Uganda had similar findings whereby in the youngest age group (15-19 years), the disparity between men and women was over four times amongst women compared to men (14.9 versus 3.6 per 1000 person-years) (53). However, in the older 20-24 years age group the disparity significantly reduced (13.8 versus 10.2 per 1000 person-years) (53). These studies demonstrate the vulnerability of young females to HIV acquisition, but the difference in risk between the genders appears to even out with advancing age.

A recent survey in Kenya revealed that sexual behaviour among young people is mostly associated with factors such as educational attainment and place of residence (16). For example, reported condom use during the last sexual encounter was higher among young people in urban areas than among their rural counterparts (16). Also, young women living in urban areas, as well as those with no education, were more likely to have greater age-disparity with their male partners (16).

Limitations of the survey findings include social desirability bias which may result from respondents failing to report sexual behaviour deemed socially inappropriate to interviewers.

Additionally, various socio-economic dynamics in low and middle-income countries such as Kenya reduce young people's ability to negotiate safer sex (14). Unemployment among young people who are in the process of transitioning to economic independence may lead to vulnerabilities motivating a proportion to engage in transactional relationships (54). This may either be to meet basic needs or to improve their social status (25). Young people from households with higher economic status may also participate in transactional sex by using material exchanges to establish intimate romantic relationships (27).

Marital or cohabiting unions are intimate relationships that may also present a higher risk of HIV acquisition (49,55). The heightened risk may be due to certain characteristics of the unions that may involve a partner practising concurrency (56). Concurrency in marriage may either be formal as in polygamy or informal as in extramarital relationships. A study in sixteen African countries found that both polygamous and extramarital relationships increased the odds of being HIV positive at the individual level (56). Nonetheless, among young people, the proportion who are in marital unions has been steadily decreasing over time as the age at marriage increases (57).

1.2.4 Association of knowledge and attitudes with sexual behaviour

In many populations, risk perception has been identified by various HIV prevention studies to precede and inform sexual conduct (58–60). A conceptual framework was used by one study in Kenya to describe the association between risk perception and sexual practices (58). In this study, socio-demographic factors such as current age, gender, ethnicity and marital status were conceptualised to act through psychosocial factors such as knowledge on HIV and attitudes towards infected people to influence an individual's risk perception and consequently sexual behaviour(58).

Despite the high HIV prevalence in SSA, the majority of young people do not consider themselves to be at high risk for HIV (61–63). This is due to relatively low levels of comprehensive knowledge on HIV and rejecting attitudes towards people living with HIV (62,63).

A recent study found that only 25.7% of young people aged 15-24 years in SSA have comprehensive knowledge on HIV (64). In this study, comprehensive knowledge was defined as having correct knowledge of at least two methods of preventing HIV acquisition through sex and rejecting two common misconceptions about HIV transmission and awareness that a healthy looking person may be infected (64). The low levels of knowledge among young people are alarming as knowledge has been shown to be a primary component of preventive behaviour in studies utilising knowledge, attitudes and practices (KAP) conceptual frameworks (65–67). Even so, it is worth noting that among this age group, knowledge of HIV has not always been shown to reduce sexual risk behaviour (13).

1.2.5 Effect of the change in sexual risk behaviour among young people on HIV prevalence

Empirical evidence from some countries in SSA notably Uganda and Zambia shows that behavioural changes among young people have contributed to reduced national HIV prevalence (68,69). A study conducted in Zambia using successive demographic and health surveys found a significant association between declining HIV prevalence and adoption of less risky sexual behaviour defined as higher levels of abstinence and condom use among young people (69).

Studies in Uganda, Zimbabwe and urban areas of Kenya and Haiti demonstrate a similar trend (70).

A deterministic mathematical model based on real data from population-based surveys was used to show an association between declines in HIV prevalence and a change of sexual behaviour assuming heterosexual transmission (70). The authors modelled HIV prevalence based on behavioural trends observed in multiple surveys and known HIV transmission dynamics to compare the predicted natural course of the epidemic with the observed prevalence changes.

They found that the prevalence declines witnessed in these countries were not possible to reproduce without sexual risk behaviour changes (70). In so doing, they excluded the possibility that the changes were as a result of the natural course of the epidemic (70). Their findings, however, are limited by biases in population sampling as they used data from women attending antenatal clinics whose results may not be generalisable to men and non-pregnant women.

1.3 Problem Statement

In Kenya, recent evidence from studies based on geographical location reveals very high HIV incidence and prevalence among young people residing in informal settlements in the capital city and around fishing areas (30,31,45,71). Few studies have evaluated the effect of knowledge, attitudes and sexual risk behaviour among young people on the spatial distribution of HIV cases on a national scale.

1.4 Justification for the Study

Spatial analysis has the benefit of detecting clustering of disease at finer geographical units than those usually reported by the DHS on a sub-national level. In 2010, upon the promulgation of a new constitution, the administrative units of Kenya changed from eight provinces to forty-seven counties (72). The prior sub-national analyses done by the DHS and other studies had been conducted at the provincial level. The ability of spatial analysis to determine the association between high-risk sexual behaviour among young people and the county-level spatial distribution of HIV infection is useful for current geographical prioritisation by county governments targeting particular population risk groups (50).

Moreover, the comparison of spatial patterns across previous time periods may illuminate trends in geographical disease progression that may not have been previously known.

1.5 Research Question

What is the relationship between sexual risk behaviour and the prevalence of HIV among young people in Kenya in 2003 and 2008?

1.6 Study Aim and Objectives

The primary aim of this research study is to determine and compare the relationship between sexual risk behaviour and the prevalence of HIV among young people in Kenya across two-time points, 2003 and 2008.

Specific objectives include;

1. To describe and compare the socio-demographic, psychosocial and sexual risk behaviour characteristics of Kenyan youth in 2003 and 2008 according to HIV infection status.
2. To determine socio-demographic, psychosocial and behavioural factors associated with the prevalence of HIV among Kenya's youth in 2003 and 2008 while accounting for spatial heterogeneity.
3. To determine the effect of sexual risk behaviour on the prevalence of HIV among Kenya's young people in 2003 and 2008 while accounting for spatial heterogeneity.

Chapter 2 Methodology

2.1 Introduction

This chapter provides a detailed description of the study setting as well as the data collection and statistical methods used in this research report.

2.2 Description of original surveys

The demographic and health surveys conducted in Kenya in 2003 and 2008 were designed to represent the entire population of the country using a representative sample of eligible men and women aged 15-49 and 15-54 years respectively. The demographic and health information of this sample of men and women were obtained at a household level.

2.2.1 Survey area

The primary surveys were conducted in Kenya, located in the Central-Eastern region of Africa. Lake Victoria, found on the Western border of the country, is the largest Lake in Africa and is shared with neighbouring Tanzania and Uganda. Kenya's coastline is situated along the Indian Ocean, and the equator crosses through the country dividing it into Northern and Southern hemispheres. The map of Kenya with both county and former provincial boundaries is depicted in figure 2.1.

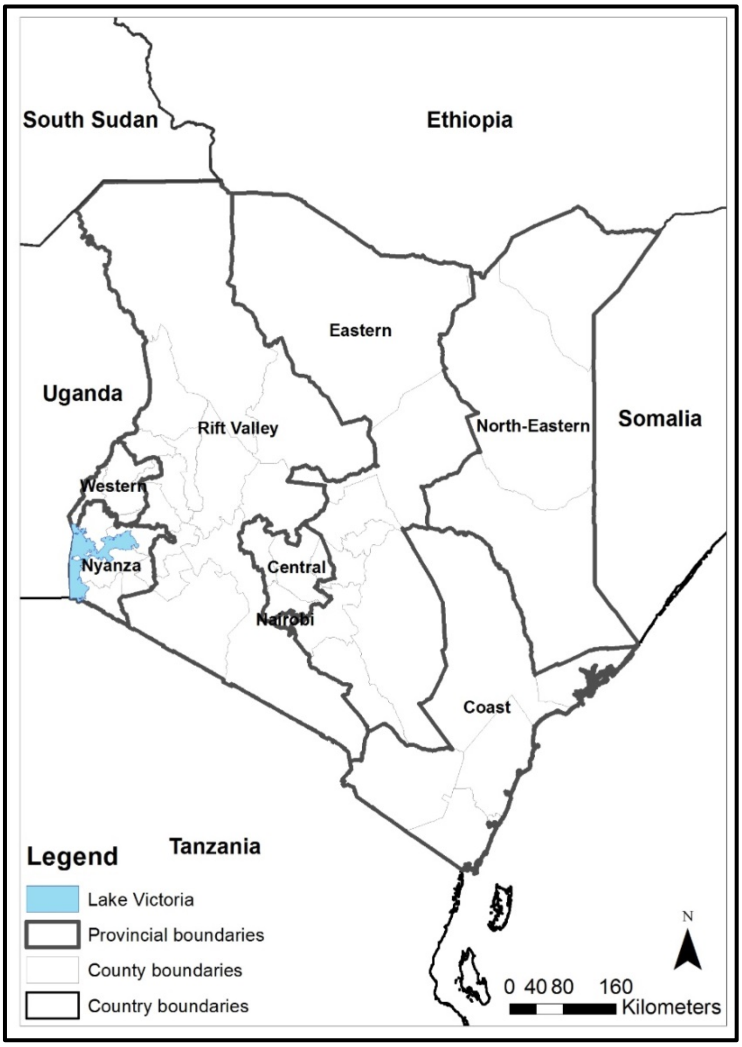


Figure 2.1 Map of Kenya and neighbouring countries

2.2.2 Survey design of the primary studies

The two DHS surveys employed a two-stage cluster sampling technique to identify participating households.

The first stage was intended to select a total of 400 clusters from the national frame of enumeration areas (7,8). Nevertheless, the 2003 and 2008 KDHS reports, state that there were 362 and 398 non-empty clusters used to randomly select houses in 2003 and 2008 respectively (7,8). The second phase included systematic sampling within the clusters to obtain a sample of households from an updated list (7,8). In the second phase, differential sampling techniques were employed whereby households in densely populated urban areas were oversampled and fewer units were obtained from the sparsely populated former North-Eastern province.

2.2.3 Data collection in original surveys

Trained interviewers administered separate questionnaires to eligible men and women in selected households. The questionnaires were modelled on the MEASURE DHS programme with adjustments to reflect the local situation in Kenya. Among respondents, those who consented were asked to provide a blood sample for HIV testing (7,8).

2.2.4 HIV testing and confirmation

Participants provided blood for HIV testing obtained from a finger prick and transported as dried blood spots (DBS) on a filter paper which was appropriately identified using a bar code and label. The samples were tested in the national reference laboratory using rapid HIV diagnostic kits (Vironostika Anti-HIV-1/2) in addition to enzyme linked immunosorbent assays (ELISA) using DADE Behring HIV-1/2 test (7,8).

Confirmation of positive results was done using an additional test (Murex HIV-1/2 MicroELISA System). At a different national laboratory, quality assurance was performed by retesting all positive samples and a 10% random sample of the negative specimen. Any discrepancies were settled using a test with higher specificity, the DNA polymerase chain reaction (PCR) (7,8).

2.3 Study population and design

This study design is repeated cross-sectional, using secondary data collected from two-stage population sample surveys in 2003 and 2008 as described in section 2.2.2.

The study population is Kenyan youth aged 15 to 24 (inclusive) years who participated in the Kenya demographic and health surveys in 2003 and 2008 and underwent HIV testing.

2.4 Power computations for the 2003 and 2008 surveys

There was no sampling for this secondary study since the primary data collection was done in the original surveys already described. However, power computations were performed using Stata for each study year separately.

The sample size of 2,650 young people in 2003 was sufficient to detect an increase in HIV prevalence among those with partners practising concurrency with a power of 94%. Adapting prevalence estimates from the 2003 KDHS report, the power computation assumed 3% population prevalence among youth. A design effect (DEFF) of 1.280 reduced the effective sample size to 2,070

Similarly, a power computation was also conducted for the 2008 sample using corresponding values adopted from the 2008/9 KDHS report. The sample size 2,857 had a power of 94% to detect an increase in HIV prevalence among those practising concurrency incorporating a design effect of 1.41

2.5 Data collection

2.5.1 Permission to obtain data

Permission to use these secondary data was obtained through a written request citing the study purpose and objectives to MEASURE DHS. The request sought authorization to access and download individual male and female demographic information, data on HIV testing as well as data with geographical attributes for both 2003 and 2008 KDHS.

MEASURE DHS granted explicit permission via email to download the relevant data from the website. The MEASURE DHS website contains data in various statistical software formats; all data were downloaded in Stata format with the exception of geographical data which were in environmental systems research Institute (ESRI) shapefile format. The geographical data provided were centroids of Kenyan DHS clusters.

The Kenyan boundaries shapefile with county coordinates was obtained from a different source, the global administrative areas website (gadm.org) (74). Explicit permission to download the shapefile from this web site was not required. However, appropriate citation is provided in the reference section (74).

2.5.2 Statistical software used for data management and analysis.

Various statistical software programs were used in this study including Stata version 14.1 from Stata Corporation LP College Station, TX. Stata was used to perform the bulk of data management and analysis. ArcMap version 10.3 from ArcGIS was used to join the county boundaries shape file with the cluster centroids from the DHS data. An attributes table of the join output was produced in text format and imported into Stata consequently merging with the demographic and HIV testing data.

Windows-based Bayesian inference using Gibbs sampling (WinBUGS), a freely available software package was used for Bayesian modelling incorporating both unstructured and structured spatial random effects (75).

R software another freely available software package was used to convert the Kenyan shape file into “Splus” format for importation into WINBUGS (76).

2.6 Data management

2.6.1 Exploration of the data

The DHS data from the two surveys were explored, retaining only variables relevant to this study. These included demographic information and variables on individual knowledge, attitudes, sexual behaviour, GPS coordinates and HIV test results.

Checks for missing values and duplicated observations were conducted using summary tables and uniquely identifying variables respectively.

2.6.2 Obtaining the final data sets used for analysis

Variables in the various datasets were renamed for ease of management. Similar variables in the male and female data sets were renamed to the same variable name to enable proper appending and merging. Merging of the DHS data routinely involves linking three main aspects between datasets; an individual, from a specific household, located within a cluster. This was done using the unique case identification, household and cluster numbers. The final datasets for each year were obtained by as illustrated in the data management flow chart in figure 2.2.

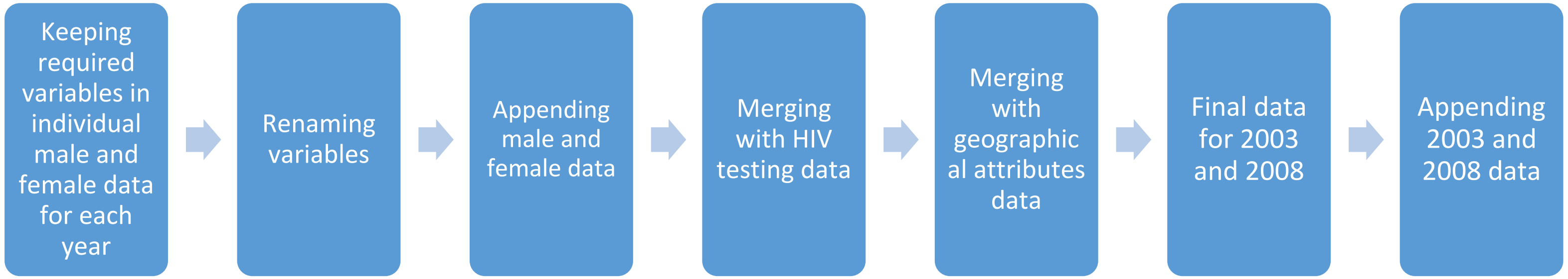


Figure 2.2–Flow diagram showing data management process

2.6.3 Sample weighting and population adjusted data

The data for each year was survey set before retaining only observations within the target age group and inclusive of both age 15 and 24 years. The variable region representing the former provinces; Nairobi, Central, Coast, Eastern, Western, Nyanza, Rift Valley and North Eastern and the variable residence representing urban or rural residence were used to determine the strata. Sample weighting was done using the HIV test results sample weights in each survey year divided by 1,000,000 (77). The samples were weighted to represent the entire population of the country. Moreover, sampling weights were also assigned to the primary sampling unit which was the cluster within which the households were randomly selected (further details are provided in Appendix A.4).

2.6.4 Creation of new variables and definition of variables

New variables were created to represent a composite score for comprehensive knowledge of HIV, attitudes towards people living with HIV and risky sex at first sexual encounter.

2.6.4.1 Creation of comprehensive knowledge score

An individual composite score on comprehensive knowledge on HIV was generated for each respondent. In this study, the definition of comprehensive knowledge included answering correctly to questions about HIV transmission and prevention plus rejecting common myths on HIV transmission for example through kissing or sharing food. Twenty questions on information related to HIV knowledge were used in the creation of the score. These questions are provided in Appendix A.6.1.

Two questions were omitted as they were not specific on whether one should avoid only unsafe blood transfusions or injections to prevent HIV. It is well known that blood transfusions with screened blood and safe medical injections are life-saving and provide minimal to no HIV risk (78,79).

Some variables were recoded to enable the combination into a composite score. For example, a correct affirmative response to a question on condom use for prevention of HIV was already coded as "1" while the right response "no" to a question on a common myth such as to prevent HIV one should avoid mosquito bites was coded as "0". Variables where correct responses were coded “0” were recoded to “1" so as to have all correct and incorrect answers coded in the same way. The overall score per individual was created by summation of correct responses for each of the 20 questions on knowledge and conversion to a percentage of the highest possible mark.

Exploratory factor analysis and confirmatory factor analysis using structural equation models were also used to create an aggregate score, and these results were used to provide sensitivity analyses by correlating with the scores obtained using the summation method already described. Results are presented in Appendix A.5.

2.6.4.2 Creation of a collective score on attitudes towards HIV infected people

A similar process as outlined in section 2.6.4.1 was used to generate an attitude score. Variables on questions probing attitudes towards people living with HIV (PLWH) were used.

In the data provided, there were two questions on attitudes towards PLWH with zero observations hence only four variables with available information were used. The questions used are provided in Appendix A.6.2. The attitude score was classified as either accepting or rejecting attitude. Accepting attitudes were coded as "1" while rejecting attitudes were coded "0".

2.6.4.3 Creation of variable on risky sex at first sexual encounter

Risky sex at first sexual encounter was defined as having sexual debut before age 15 or failure to use a condom at sexual debut or engaging in intercourse at first meeting. The new variable was consequently coded "1" for respondents who reported any of the aforementioned characteristics and conversely coded "0" for those who were yet to initiate sexual activity or had initiated sex after age 15 years or used a condom at first sex.

2.6.4.4 Recoding and categorisation of other behavioural variables

High-risk sexual behaviour variables were defined by recoding variables on respondents' sexual conduct. Risky behaviour was coded as "1" if a respondent reported a cross-generational sexual partner (more than ten years older) or engaged in transactional sex (gifts/money given or received for sex) or reported that their sexual partner had other concurrent partners. The number of sexual partners within the previous twelve-month period was retained as a continuous variable. Risky sex was coded “0” for respondents who had not iniatiated sexual activity for all risky sex variables.

Categorisation of age at sexual debut into four categories was done using the definition of early sexual debut described in section 1.2.1 of Chapter 1.

The four categories denote those who had not had sex, sexually experienced initiating sex before age 15 years and after 15 years and lastly, those who had sex at their first union age not specified as in the original data.

2.6.4.5 Categorisation of socio-demographic and psychosocial variables

Educational attainment was re-categorised into three categories to reflect no education, primary and secondary or higher education. The original grouping of other socio-demographic variables used in this study was maintained as in the original data.

The generated percentage knowledge and attitude scores were categorised into binary variables with 50% being the cutoff for each person's comprehensive knowledge score. This categorisation was done for descriptive analysis only. Respondents who scored less than 50% were classified as not having comprehensive knowledge and those with scores 50% or higher classified as having comprehensive knowledge. The same was done for accepting and rejecting attitude scores. These scores were maintained as continuous percentages in the modelling analysis.

2.6.5 Multiple imputations of missing values

The summary tables for variables with missing values revealed a significant proportion of missing values of some of the sexual risk behaviour variables. These were variables representing having a cross-generational partner, engaging in transactional sex and having a partner who practices concurrency.

Stata was used to determine the pattern of “missingness” among variables of interest. The missing values were assumed to be missing at random (MAR) that is the probability that a variable is missing depends only on available information (80).

As a result, logistic regressions using multivariate imputation by chained equations (MICE) methods were used to impute the missing binary sexual behaviour variables (81). The logistic regression imputation model included other observed variables such as age, gender, wealth index as well as other sexual risk behaviour and specified the creation of 20 additional imputed datasets.

After imputations, the proportions of each imputed variable were compared against the original observed variable without imputed values for all the three variables of interest. The match with the closest similarity in summary statistics was retained for further use in the modelling.

2.7 The measures and variables of the first objective

The first objective was to describe and compare the socio-demographic, psychosocial and sexual risk behaviour of Kenyan youth in 2003 and 2008 according to HIV infection status.

The outcome variable was HIV test results which were measured as described in section 2.2.4. Other variables described included the socio-demographic variables such as age, gender, education and residence. Psychosocial variables included comprehensive knowledge and attitudes towards PLWH while sexual behaviour variables described were cross-generational sexual partnerships, transactional sex, number of sexual partners, sexual debut and condom use at sexual debut.

2.7.1 Descriptive analysis of the outcome and explanatory variables

Descriptive analysis was performed for selected socio-demographic and sexual behaviour variables while accounting for the survey design and calculated sample weights. Comparison of proportions between the two survey years was done using a corrected design based F statistic (Rao-Scott adjustment) and simple survey linear regression was used to compare means.

Descriptive tables and maps of HIV prevalence by county were also produced and are reported in chapter 3.

2.8 The dependent and explanatory variables of the second objective

The second objective was to determine socio-demographic, psychosocial and behavioural factors associated with the HIV prevalence among Kenya's youth in 2003 and 2008 while accounting for spatial heterogeneity.

The dependent variable was the HIV test results as described in section 2.2.4.

Explanatory variables included the Global Positioning System (GPS) coordinates of the respondents and geo-coded shape files for Kenya. Psychosocial factors; comprehensive knowledge of HIV and attitudes towards PLWH. Socio-demographic variables; age, gender, education, wealth index and residence. The sexual behaviour variables were transactional sex, cross-generational partner, risky sex at first sexual encounter, the number of partners in previous twelve months and having a partner who practices concurrency.

2.8.1 Univariable and multivariable analysis

Spatial analysis was done by fitting multilevel survey logistic regression models adjusting for county random effects. Firstly, for each study year, univariable survey logistic regression models with HIV status as the outcome were fitted for each of the explanatory variables mentioned in section 2.8.

Secondly, multivariable survey logistic models were fitted using the variables that were significant during univariable analysis. The best fitting survey logistic model was then specified as a multilevel model allowing for random intercepts across counties and assuming fixed effects of the specified covariates. Taking the survey design into account during modelling allows inferences to be made at an individual-level, given that persons residing in the same cluster or household may not be entirely independent. Separate models were constituted for each year of analysis and results are reported in Chapter 3. The datasets were then appended, and a final parsimonious model fitted incorporating the year of the survey as a covariate was constituted.

Some variables that were not significant in univariable analysis were retained in the model as variables of interest, for example, the variable representing having a cross-generational partnership. The respondents' current age and the number of sexual partners in past 12 months were included as continuous predictors in the model while other variables were denoted as categorical factors.

In the separate year models, improvements to model fit were tested by stepwise addition of other variables and possible interaction terms to the survey model then testing using the seemingly unrelated estimation procedures (suest test).

The final model for the year 2003, included the variable representing the wealth index while it was omitted from the model for 2008. Other variables such as comprehensive knowledge scores and attitude scores that were not significant in univariable analysis and did not improve model fit using the seemingly unrelated estimation were omitted so as to have a final parsimonious model.

The final models in both years were mixed effects models which had both fixed and random covariates. The fixed covariates were age, gender, residence, educational status, transactional sex, cross-generational partner, risky first sexual encounter, the number of sexual partners in previous 12 months, and reporting a partner who had other concurrent partners. The random variable county was treated in the models as the unstructured random effects.

2.8.1.1 Model goodness of fit

The overall model goodness of fit was also assessed using the seemingly unrelated estimation testing cross-model hypothesis. The null hypothesis of the test was that the coefficients of the outcome compared in two models were the same. Failure to reject the null hypothesis suggested that the model was a good fit to the data (82). Details of the Stata commands used, and the test results are provided in Appendix A.7.

2.8.1.2 Mapping the predicted estimates after fitting the model

The mean individual probabilities of the outcome and the standard errors were predicted using the predict command after fitting the models for each year then collapsed by county. Stata packages (spmap and shp2data) were then used to map the average predicted individual probabilities of infection collapsed by county using the Kenyan county coordinates map (83).

2.8.2 Bayesian modelling

The final maximum likelihood regression model in Stata was replicated in WinBUGS, a software package that can be used for Bayesian estimation of parameters (84).

The Bayesian methodology involves the use of priors in updating the parameters derived from the likelihood. The resulting updated product of the likelihood, the priors and hyper-priors (priors of the priors) is known as the posteriors distribution.

$$\mathbf{Posteriors}(\mathbf{parameters\ given\ the\ data})$$
$$= \mathbf{Likelihood}\ (\ \mathbf{data\ given\ parameters})\mathbf{x\ Priors}\ (\mathbf{distributions\ of\ parameters})$$

The posterior distribution, therefore, refers to the conditional distribution of parameters in the simulations after observing the actual data (84).

Firstly, the data was converted within Stata to BUGS format before importation into WinBUGS. The Bayesian model with fixed covariates was first specified and updated with only unstructured spatial random effects to compare to the Stata model. Consequently, the same model incorporating both unstructured and structured spatial random effects was also compiled.

The structured spatial component was included as an adjacency matrix generated using the Kenyan map with county boundary regions added to the data. A conditional autoregressive (CAR) model utilising Tobler’s first law of geography which states that everything is related to everything else, but near things are more related than distant things was used to account for spatial autocorrelation (85). This was the basis used to adjust for the structured spatial random effects, in the convolution model (model with structured and unstructured effects). Ignoring spatial autocorrelation leads to exaggerated precision errors as standard errors appear smaller than in reality hence the spatial model was the preferred final model over the model with only unstructured county random effects.

The Bayesian spatial model was specified as a full joint distribution with known priors for the beta coefficients, structured and unstructured random effects. The priors used are provided in Appendix A.8. The model was compiled with two chains because the convergence of simulations can be checked by running more than one chain (84). The chains contained specified initial values also provided in Appendix A.8. The model was updated using Markov Chain Monte Carlo methods (MCMC) after compilation (84).

The first 5,000 iterations were treated as parameter adjustments (burning in) followed by a further 15,000 iterations while thinning every five iterations to minimise the autocorrelation of the subsequent iterations. The resulting 3,000 simulations were used for parameter estimation after establishing convergence, as detailed in section 2.8.2.1. The parameter summary statistics with the corresponding 95% credible intervals were estimated from sampling the posterior distribution.

2.8.2.1 Monitoring convergence and model diagnostics

The convergence of simulations was monitored using trace and autocorrelation plots for the specified model parameters. Examining the different parameter plots, it was reasonable to conclude that convergence was achieved since the two chains appearing in different colours seemed to be overlapping from the first iteration after burning. The trace and autocorrelation plots are provided in Appendix A.9.

Model diagnostics were performed by comparing the Deviance Information Criterion (DIC) of the unstructured and structured model. The model with the best fit is that with the smallest DIC (86). The DIC of both models are included in Table 3.7 in chapter 3.

2.8.2.2 Producing maps after Bayesian modelling

The mapping tool in WinBUGS was used to create a choropleth map of Kenya illustrating the structured spatial random effects by county. The individual mean posterior probabilities of the outcome were imported into Stata then collapsed by county over the sum of observed cases.

Mapping of the obtained individual posterior probabilities averaged by county as well as the unstructured and structured county random effects was done using the Kenyan county coordinates map in Stata.

2.9 The outcome and explanatory variables of the third objective

The third objective was to determine the effect of high-risk sexual behaviour on the prevalence of HIV among Kenya's young people in 2003 and 2008 while accounting for spatial heterogeneity.

The main outcome variable was HIV status while explanatory variables included the comprehensive knowledge and attitude scores, high-risk sexual behaviour variables and the variable representing the county.

2.9.1 Generalised structural equation modelling

The conceptual framework depicted in figure 2.3, was the basis of the generalised multilevel structural equation model (GSEM) fitted in Stata using the GSEM builder. GSEM has an advantage over classical regression modelling in situations where individual variables of interest cannot be measured perfectly (87).

This is because the model takes into account the measurement error in the observed continuous variables during estimation. Moreover, GSEM allows for modelling of both direct and indirect effects of the covariates on the outcome.

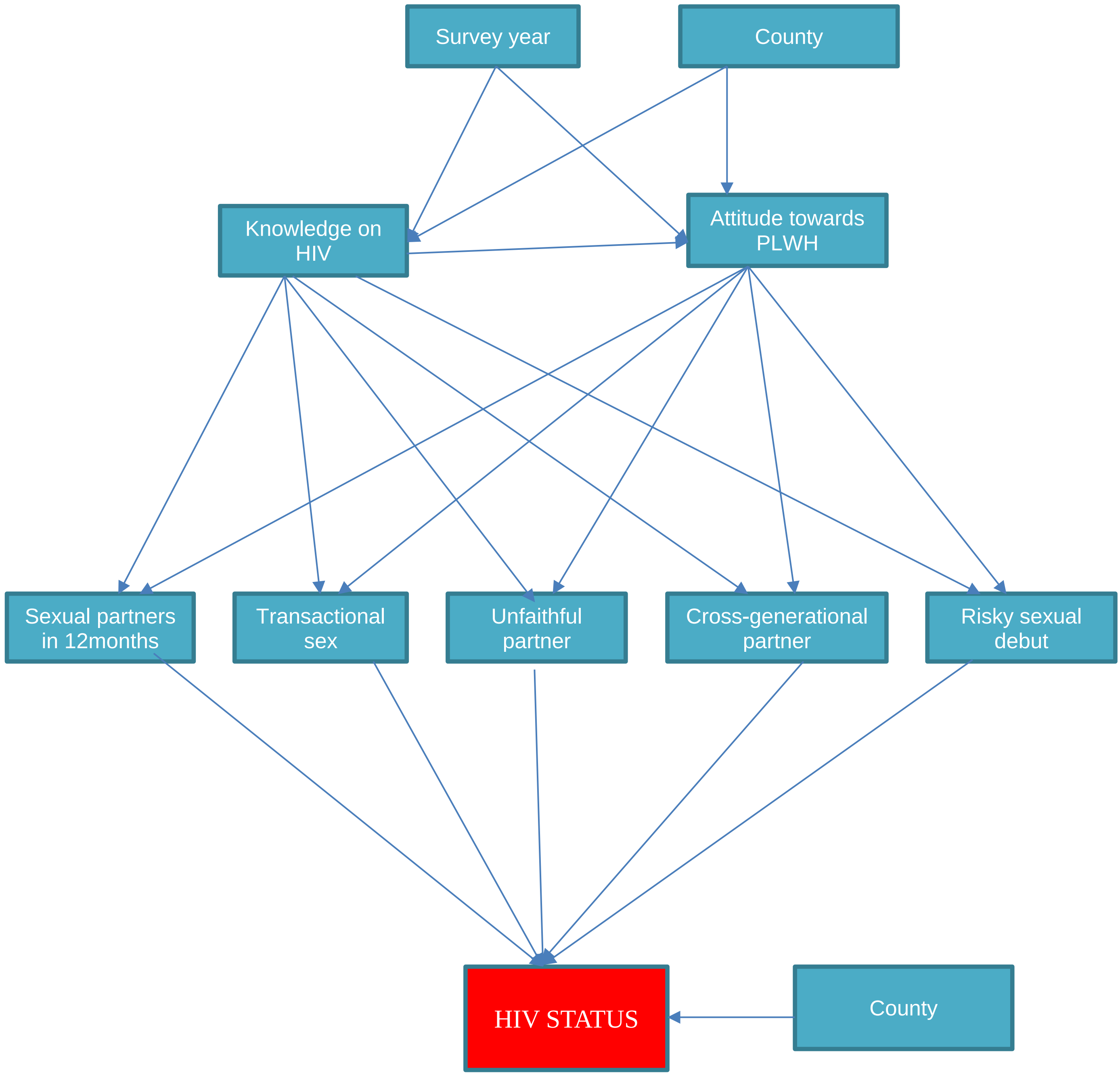


Figure 2.3: Diagram of knowledge, attitudes and practices (KAP) conceptual

In the GSEM, comprehensive knowledge on HIV represented as a continuous percentage score (already described in section 2.6.4.1) was conceptualised to directly affect attitudes towards PLWH while indirectly affecting HIV status through the sexual behaviour variables. Attitudes towards PLWH represented as a continuous percentage score (described in section 2.6.4.2) was also modelled to affect HIV status through the sexual behaviour variables.

The sexual behaviour variables of interest; transactional sex, cross-generational partner, the number of sexual partners in previous 12 months, risky sex at first sexual encounter and reporting a partner who had other concurrent partners were each modelled to have direct effects on the outcome, HIV status. The county, denoted as a multilevel variable (double ring) in the GSEM structure in figure 3.8, was used to adjust for random effects representing the heterogeneity among counties in the overall outcome. The model also included the year of the survey as a dichotomous exogenous variable representing either 2003 or 2008 modelled to influence the other explanatory variables directly.

The final graphical representation of the model was reported omitting pathways that were not significant at the 5% level of significance. Pathway analyses of direct, indirect and total effects of covariates on the outcome were conducted and reported in a tabular format as coefficients. The indirect effects of covariates on the outcome were obtained by summing coefficients along all paths from the specific variable to the main outcome. The total effect was the cumulative direct and indirect effects of each respective variable. The details are provided in Appendix B.1.

2.10 Ethical Considerations

During the primary surveys, the interviewers obtained informed consent from respondents before HIV testing. Explicit permission to use the MEASURE DHS data was obtained as highlighted in section 2.5.1. The data downloaded from MEASURE DHS website did not contain any personal identifiable information. Also, the provided GPS coordinates were estimated centres of clusters as opposed to individual household locations. MEASURE DHS further anonymised the GPS locations through a random displacement of coordinates which is restricted so that points stay within the country and the second administrative region.

Furthermore, specific ethical approval for this study was obtained from the University of the Witwatersrand Human Research Ethics Committee. The ethics clearance certificate number is M1611100 (copy of certificate attached in Appendix B.2).

Chapter 3 Results

3.1 Introduction

This chapter outlines the results obtained from the analyses for each of the objectives described in chapter 2. Results are represented as summary narratives as well as in tabular and graphical format.

3.2 Results of the descriptive analysis

3.2.1 National HIV prevalence among the youth and the general population

The total numbers of survey respondents aged 15-24 years with complete information on HIV test results were 2,650 and 2,857 in 2003 and 2008 respectively. The flow diagram in figure 3.1 shows how these final samples were obtained.

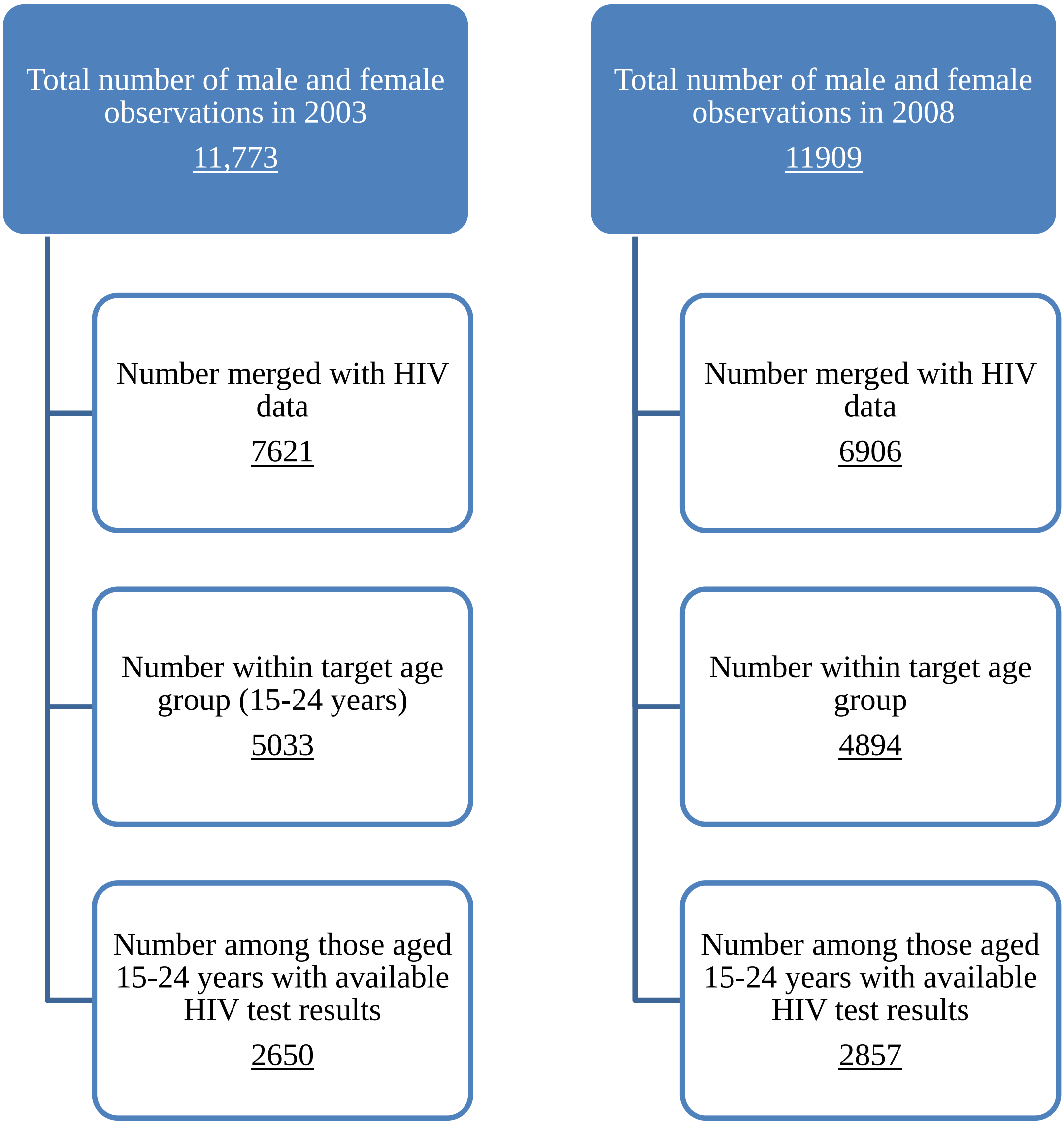


Figure 3.1 Flow diagram showing how final samples were obtained

The HIV prevalence among young people aged 15-24 years for the weighted samples decreased from 3.61% in 2003 to 2.88% in 2008, but this difference was not statistically significant. Results are illustrated in figure 3.2 and table 3.2 and 3.3.

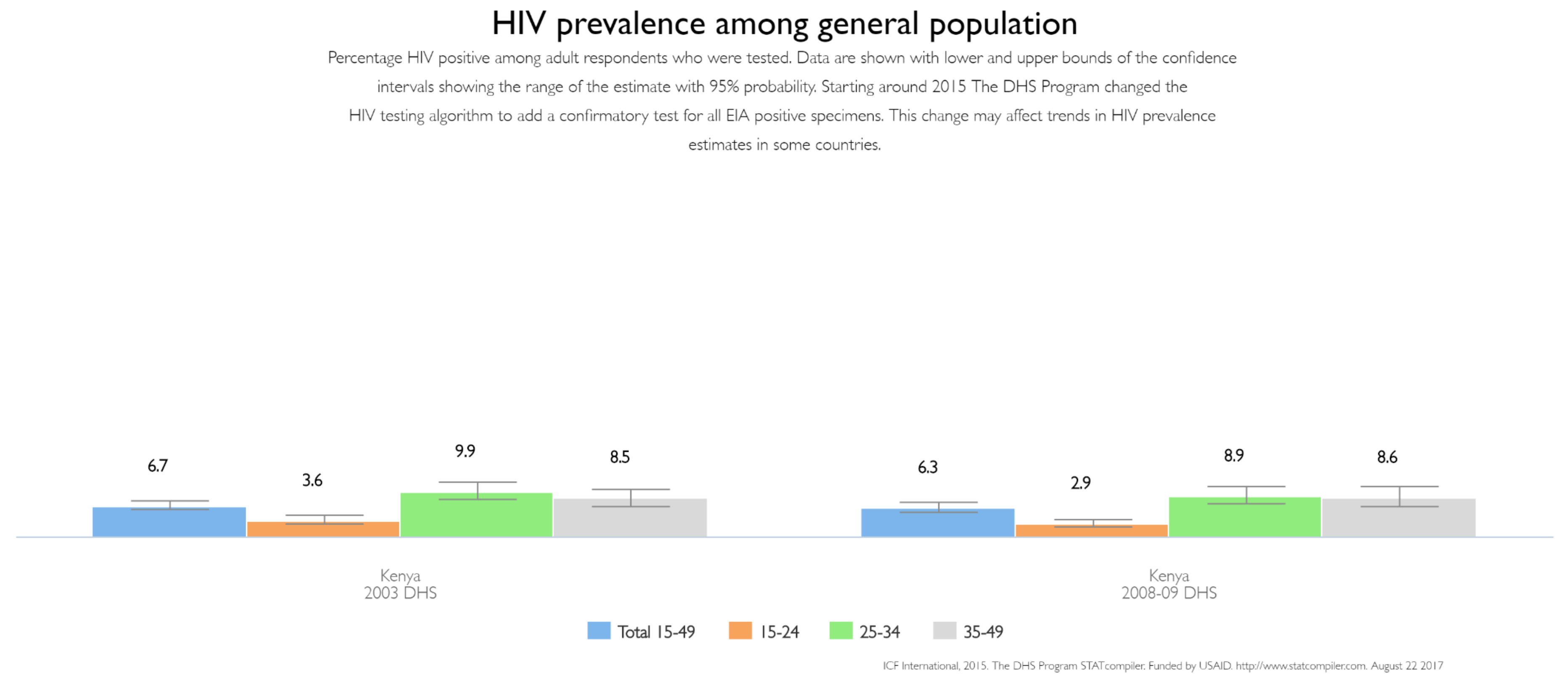


Figure 3.2 HIV prevalence among the Kenyan general population including youth in 2003 and 2008

3.2.2 County level HIV prevalence among youth

Table 3.1 shows the HIV prevalence among young Kenyan people by county in 2003 and 2008 for counties with available HIV test results. Additionally, maps depicting the county HIV prevalence in 2003 and 2008 are illustrated in figure 3.3 and figure 3.4.

Table 3.1: County level HIV prevalence among Kenyan youth in 2003 and 2008

County HIV prevalence among Kenyan youth in 2003 and 2008				
			Percentage prevalence	
County code	**County Name**	**Former Province**	**2003**	**2008**
1	Mombasa	Coast	2.8	5.56
2	Kwale	Coast	3.13	1.92
3	Kilifi	Coast	2.3	3.03
4	Tana River	Coast	9.52	*
5	Lamu	Coast	*	*
6	Taita–Taveta	Coast	3.85	*
7	Garissa	North Eastern	*	*
8	Wajir	North Eastern	*	*
9	Mandera	North Eastern	*	3.57
10	Marsabit	Eastern	*	*
11	Isiolo	Eastern	*	*
12	Meru	Eastern	*	2.9
13	Tharaka-Nithi	Eastern	*	2.7
14	Embu	Eastern	*	1.75
15	Kitui	Eastern	1.64	*
16	Machakos	Eastern	3.08	2.33
17	Makueni	Eastern	1.64	3.13
18	Nyandarua	Central	2.9	2.99
19	Nyeri	Central	*	*
20	Kirinyaga	Central	5.13	2.56
21	Murang'a	Central	4.48	1.41
22	Kiambu	Central	6.34	5.88
23	Turkana	Rift Valley	*	*
24	West Pokot	Rift Valley	*	*
25	Samburu	Rift Valley	10	13.33
26	Trans-Nzoia	Rift Valley	*	3.77
27	Uasin Gishu	Rift Valley	7.02	*
28	Elgeyo-Marakwet	Rift Valley	*	*
29	Nandi	Rift Valley	*	*

County HIV prevalence among Kenyan youth in 2003 and 2008

County code	**County Name**	**Former Province**	**Percentage prevalence**	
			2003	**2008**
30	Baringo	Rift Valley	*	*
31	Laikipia	Rift Valley	*	*
32	Nakuru	Rift Valley	5.26	*
33	Narok	Rift Valley	3.7	5.71
34	Kajiado	Rift Valley	0	9.52
35	Kericho	Rift Valley	4	3.33
36	Bomet	Rift Valley	2.44	*
37	Kakamega	Western	4.79	0.68
38	Vihiga	Western	1.61	2.86
39	Bungoma	Western	0.83	1.19
40	Busia	Western	1.52	6.9
41	Siaya	Nyanza	8.22	12.66
42	Kisumu	Nyanza	14.52	10.32
43	Homa Bay	Nyanza	13.43	15.79
44	Migori	Nyanza	10.71	1.96
45	Kisii	Nyanza	3.7	3.16
46	Nyamira	Nyanza	9.76	2.63
47	Nairobi	Nairobi	3.89	3.36

(*) Missing HIV test results

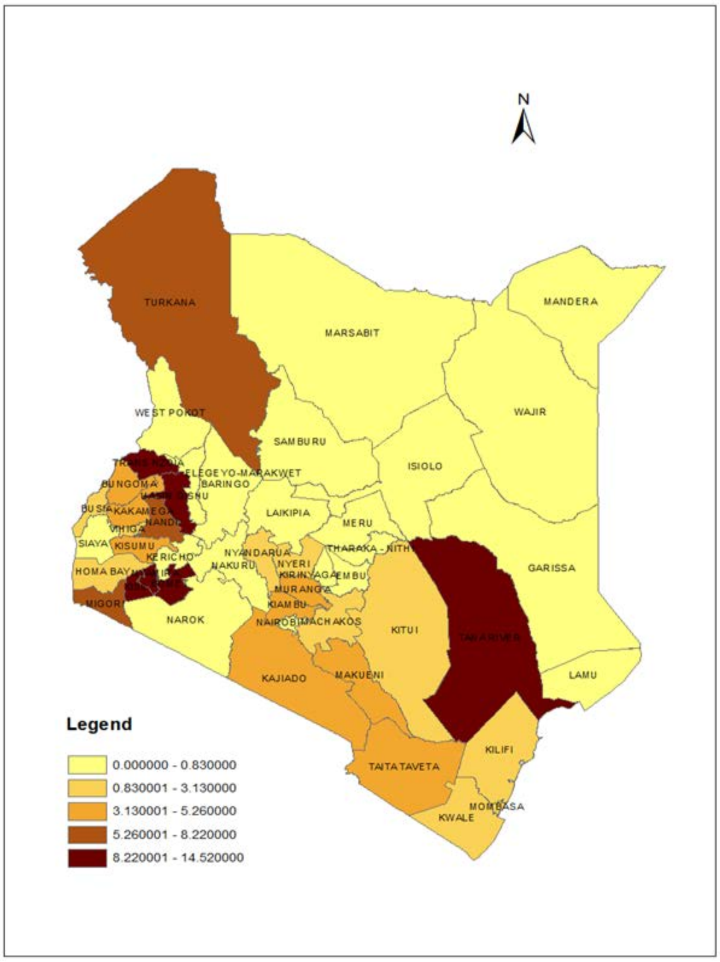


Figure 3.3 Observed HIV prevalence by county among Kenyan youth in 2003

Figure 3.4 shows the HIV prevalence among Kenyan youth in 2008. Counties in the Western region of Kenya had the highest observed prevalence.

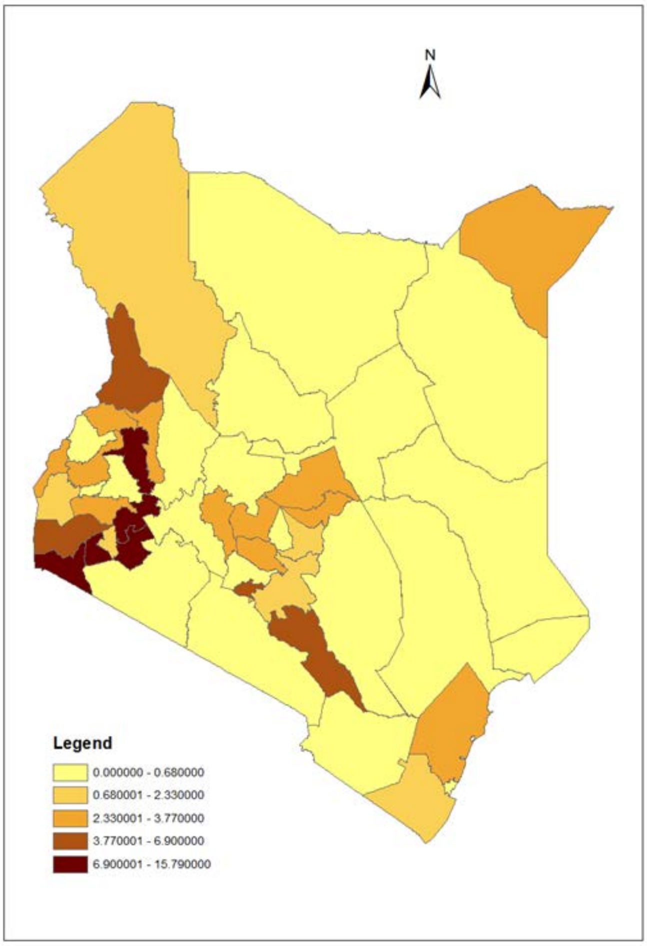


Figure 3.4 Observed HIV prevalence by county among Kenyan youth in 2008

3.2.3 Summary of missing and imputed values of sexual behaviour variables

The complete, missing, and imputed values of some variables of interest, particularly the variables on sexual risk behaviour which had many missing values are represented in table 3.2.

Table 3.2: Complete, missing and imputed values of sexual risk behaviour variables

	2003				2008			
Variables	Complete	Missing	Imputed	Total	Complete	Missing	Imputed	Total
Transactional sex	2494	156	156	2650	2606	251	251	2857
Cross generational partner	1057	1593	1593	2650	1650	1207	1207	2857
Partner with concurrent partners	1326	1324	1324	2650	1451	1406	1406	2857

3.2.4 Description of socio-demographic, psychosocial and sexual risk behaviour variables by HIV status

In both years of analyses, there was overwhelming evidence of an age and gender difference in HIV prevalence as illustrated in tables 3.3 and 3.4 .The mean age was higher among HIV positive survey respondents, and the majority were female in both 2003 and 2008. Most of the surveyed youth were rural residents and had attained only primary level education.

A significant majority of young people were already sexually active with the greatest proportion of infected youth reporting sexual initiation after age 15 years. Although the majority did not use condoms at sexual debut, there was insufficient evidence that the proportions infected differed with those who reported condom use at sexual debut. A small proportion reported engaging in transactional sex in both years. In 2008 unlike in 2003, there was a significant difference in HIV status among youth who engaged in transactional sex and those who did not engage.

Notably, among those who were HIV positive in 2003, none reported having a cross-generational sexual partner. In both study years, there was a significant difference in HIV status among those who said they had a partner with other concurrent sexual partners and those who did not.

Comprehensive knowledge on HIV and accepting attitudes towards PLWH increased significantly between 2003 and 2008. The knowledge score representing comprehensive awareness of HIV increased from 42% in 2003 to 78% in 2008. There were also significant changes in sexual behaviour among young people in the two years. More young people reported less risky sexual behaviour in 2008 compared to 2003 except cross-generational partnerships which the proportion reporting a cross-generational partner increased.

Table 3.3: Socio-demographic, psychosocial and behavioural characteristics by HIV status among Kenyan youth in 2003

Variables		**HIV Positive**	**HIV Negative**
	category	**Weighted %**	**Weighted %**
HIV test results		3.61	96.39
Age (years)	mean ±SE	21.15 ±0.27	19.14 ±0.07
Gender	male	16.82	50.12
	female	83.18	49.88
Education	No education	4.45	5.083
	Primary Education	74.53	65.04
	Secondary or higher	21.02	29.87
Residence	Urban	35.22	23.82
	Rural	64.78	76.18
Comp HIV Knowledge	No	59.36	57.79
	Yes	40.64	42.21
Attitudes towards PLWHA	Rejecting	17.84	19.61
	Accepting	82.16	80.39
Sex debut	Not had sex	12.82	35.51

Variables		HIV Positive	HIV Negative
	category	**Weighted %**	**Weighted %**
	Debut before age 15	30.25	21.22
	Debut after age 15	53.17	39.53
	Sex at first union	3.77	3.74
Condom use at sexual debut	No	86.64	86.42
	Yes	13.36	13.58
Cross generational partner	About same age	23.3	38.21
	less than 10 years older	76.7	57.31
	10 or more years older	0	4.48
No. of partners in previous 12 months	Mean ±SE	1.18 ±0.02	1.03 ±0.03
Transactional sex	No	93.76	91.40
	Yes	6.24	8.60
Partner has other partners	No	24.1	60.32
	Yes	75.9	39.68

Table 3.4: Socio-demographic, psychosocial and behavioural characteristics by HIV status among Kenyan youth in 2008

Socio-demographic characteristics		HIV Positive	HIV Negative
		Weighted %	**Weighted %**
HIV test results		2.88	97.12
Age(years)	Mean ±SE	20.73 ±0.32	19.26 ±0.08
Gender	Male	17.57	48.7
	Female	82.43	51.3
Education	No education	5.40	3.64
	Primary	66.94	57.49
	Secondary or higher	27.66	38.88
Residence	Urban	22.31	21.06
	Rural	77.69	78.94
Comprehensive HIV knowledge	No	19.61	22.28
	Yes	80.39	77.72
Attitudes towards PLWHA	Rejecting	7.193	10.99
	Accepting	92.81	89.01
Sex Debut	Not had sex	12.23	39.28

Socio-demographic characteristics		HIV Positive	HIV Negative
		Weighted %	Weighted %
	Debut before age 15	23.65	14.91
	Debut after age 15	62.35	43.77
	Sex at first union	1.78	2.05
Condom use at sexual debut	No	72.01	73.88
	Yes	27.99	26.12
Transactional Sex	No	92.09	96.69
	Yes	7.91	3.31
Cross generational partner	10 or more years older	21.74	23.81
	less than 10 years older	77.23	74.77
	older, unsure how much	1.027	1.424
No. of partners in previous 12 months	Mean ±SE	1.15 ±0.02	1.09 ±0.05
Partner has other partners	No	53.94	80.81
	Yes	46.06	19.19

Table 3.5: Comparison of socio-demographic, psychosocial and behavioural factors among Kenyan youth in 2003 and 2008

Variables		2003	2008	F- statistic, P value
		Weighted %	Weighted %	
HIV test results	Positive	3.61	2.88	F=1.65, 0.1989
	Negative	96.39	97.12	
Age(years)	mean ±SE	19.21 ±0.07	19.30 ±0.08	0.393*
Gender	male	48.92	47.81	F=0.46,0.4991
	female	51.08	52.19	
Education	No education	5.06	3.69	F=8.62,0.0002
	Primary	65.39	57.76	
	Secondary or higher	29.55	38.56	
Residence	Urban	24.23	21.1	F=0.74,0.3887
	Rural	75.77	78.9	
Comprehensive HIV knowledge	No	57.85	22.2	F=307.27, <0.0001
	Yes	42.15	77.8	
Attitudes towards PLWHA	Rejecting	19.54	10.88	F=41.11, <0.0001
	Accepting	80.46	89.12	

Variables		2003 Weighted %	2008 Weighted %	F- statistic, P value
Sex Debut	Not had sex	34.7	38.5	F= 9.63, <0.0001
	Debut before age 15	21.54	15.16	
	Debut after age 15	40.02	44.3	
	Sex at first union	3.74	2.04	
Condom use at sexual debut	No	86.43	73.8	F=40.96, <0.0001
	Yes	13.57	26.2	
Transactional Sex	No	91.51	96.48	F=18.74, <0.0001
	Yes	8.49	3.52	
Cross generational partner	No	95.71	74.98	F=27.26, <0.0001
	Yes	4.29	25.02	
No. of partners in previous 12 months	Mean ±SE	1.27 ±0 .10	1.14 ±0.02	0.201*
Partner has other partners	No	57.96	79.13	F=31.25, <0.0001
	Yes	42.04	20.87	

(*) P-value obtained from linear survey regression

3.3 Results of univariable and multivariable regression modelling

This section highlights the results of modelling to determine the factors that were associated with the distribution of HIV.

3.3.1 Factors associated with the HIV distribution among youth in 2003

In 2003, gender had the greatest effect on HIV status during both univariable and multivariable analysis adjusting for unstructured county random effects. The results are represented in table 3.6. During multivariable analysis, the likelihood of being HIV infected was more than three times among females as compared to males (AOR 3.47; 95%CI 1.97- 6.10). Additionally, for every one year rise in age, the likelihood of being infected significantly increased by 23% (AOR 1.23 95%CI 1.14-1.34) adjusting for other specified demographic factors and risky sexual behaviour.

There was some evidence during univariable analysis of an association between the level of educational attainment and HIV status. Having secondary or higher education level was significantly protective against HIV infection compared to those with only primary education (OR 0.56; 0.34-0.93). This effect was also observed during multivariable analysis (AOR 0.45; 0.26-0.77).

There was also some evidence that engaging in risky sex at first sexual encounter (AOR 1.62; 95%CI 0.95-2.74 (P=0.074) and having a partner with other concurrent partners (AOR1.53; 95%CI 0.97-2.40 (P=0.064) increased the likelihood of being HIV positive while keeping other factors constant. Since there were no HIV positive respondents reporting having a cross-generational partner, the variable was not a predictor of HIV infection in 2003.

Table 3.6: Factors associated with HIV infection among Kenyan youth in 2003

2003 Multilevel Logistic Regression results			
Variables		**Unadjusted OR; 95%CI (P value)**	**Adjusted OR; 95%CI (P value)**
Age		1.26; 1.16-1.36 (<0.0001)	1.23; 1.14-1.34 (<0.0001)
Gender	Male	1.00	1.00
	Female	4.30; 2.55- 7.23 <0.0001)	3.47; 1.97- 6.10 (<0.0001)
Education	Primary education	1.00	1.00
	No education	0.94; 0.38- 2.33 (0.889)	0.71; 0.28-1.81 (0.478)
	Secondary or higher	0.56; 0.34-0.93 (0.025)	0.45; 0.26-0.77 (0.003)
Residence	Rural	1.00	1.00
	Urban	2.08; 1.29-3.35 (0.003)	1.49; 0.78-2.84 (0.226)
Wealth index	Poorest	1.00	1.00
	Poorer	1.28; 0.58-2.83 (0.549)	1.24; 0.55-2.84 (0.604)
	middle	0.43; 0.16-1.21 (0.109)	0.44; 0.15-1.27 (0.129)
	Richer	1.74; 0.82-3.70 (0.148)	1.82; 0.82-4.05 (0.142)
	Richest	2.61; 1.24-5.49 (0.011)	2.23; 0.89-5.61 (0.088)
Comprehensive Knowledge of HIV		1.00; 0.98-1.01 (0.740)	*
Attitudes towards PLWH		1.00; 0.99- 1.00 (0.903)	*
Transactional sex	No	1.00	1.00
	Yes	0.62; 0.27-1.45 (0.273)	0.61; 0.23-1.59 (0.311)
Number of sex partners in 12 months		1.02; 0.97-1.07 (0.478)	1.02; 0.94-1.11 (0.627)
Risky sex at debut	No	1.00	1.00
	Yes	1.33; 0.82-2.16 (0.251)	1.62; 0.95-2.74 (0.074)
Partner has other partners	No	1.00	1.00
	Yes	2.27; 1.50-3.44 (<0.0001)	1.53; 0.97-2.40 (0.064)

***** Not included in multivariable model

The predicted individual probability of being HIV infected fitted after running the regression model indicated hot spots of infection in several counties indicated by the red colour as shown in figure 3.5

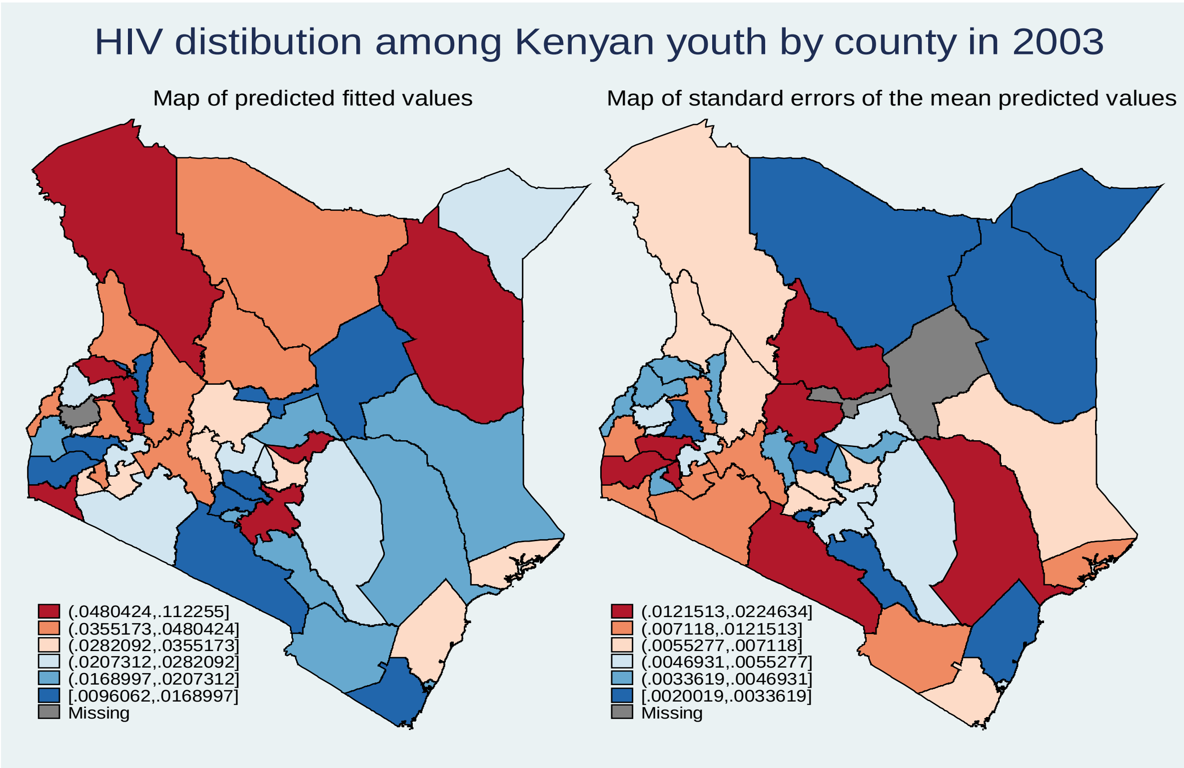


Figure 3.5 Maps showing the mean predicted fitted probability of infection and standard errors of the means by county in 2003

3.3.2 Factors associated with the HIV distribution in 2008

The detailed results of the univariable and multivariable logistic regression analysis are represented in table 3.7.

In contrast to 2003, behavioural factors were more significant predictors of HIV infection in 2008. Respondents who engaged in transactional sex of any kind either giving or receiving gifts or money for sex had the greatest likelihood of being HIV positive. Their likelihood of being infected was four and a half times that of those who did not engage in transactional sex while adjusting for unstructured county random effects and keeping other factors constant (AOR 4.53; 95%CI 2.14-9.58).Other significant behavioural determinants of HIV infection were having a partner with concurrent partners (AOR 2.50; 95%CI 1.60-3.90) and engaging in risky sex at sexual debut (AOR 1.71; 1.01-2.92). Having a cross-generational partner significantly increased the odds of being HIV positive during univariable analysis however while adjusting for other factors the increased odds were only marginally significant.

Similarly, as observed in 2003, demographic factors such as age and gender were also significantly associated with HIV status. The likelihood of being infected increased significantly by 19% for every year the respondents aged (AOR 1.19; 95%CI 1.09-1.30) and females were more than three times likely to be positive as compared to males (AOR 3.38; 95%CI 1.91-5.96). During univariable analysis controlling only for unstructured county effects, there was evidence that respondents' residence was associated with HIV status with urban dwellers more likely to be HIV infected (OR 1.84; 95%CI 1.11-3.04). However, while adjusting for other factors, the residence was only marginally associated with HIV status (AOR 1.61; 0.94-2.76, P=0.08).

Table 3.7: Factors associated with HIV infection among Kenyan youth in 2008

2008 Multilevel Logistic Regression Results			
Variables		Unadjusted OR; 95%CI (P value)	Adjusted OR; 95%CI (P value)
Age		1.22; 1.13-1.32 (<0.0001)	1.19; 1.09-1.30 (<0.0001)
Gender	male	1.00	1.00
	female	4.05; 2.37- 6.93 <0.0001)	3.38; 1.91-5.96 (<0.0001)
Education	Primary education	1.00	1.00
	No education	2.04; 0.86- 4.82 (0.104)	1.14; 0.46-2.81 (0.783)
	Secondary or higher	0.47; 0.29-0.78 (0.004)	0.47; 0.27-0.80 (0.005)
Residence	Rural	1.00	1.00
	Urban	1.84; 1.11-3.04 (0.017)	1.61; 0.94-2.76 (0.08)
Comprehensive Knowledge of HIV	No	1.00	
	Yes	0.98; 0.97-1.00 (0.068)	*
Attitudes towards PLWHA	Rejecting	1.00	
	Accepting	0.99; 0.99- 1.00 (0.090)	*
Transactional sex	No	1.00	1.00
	Yes	4.54; 2.22-9.28 (0.000)	4.53; 2.14-9.58 (<0.0001)
Cross generational partner	No	1.00	1.00
	Yes	1.74; 1.18-2.74 (0.007)	1.54; 0.97-2.44 (0.065)
Number of sex partners in 12 months		1.36; 1.13-1.64 (0.001)	1.23; 0.92-1.63 (0.157)
Risky sex at debut	No	1.00	1.00
	Yes	1.92; 1.19-3.12 (0.008)	1.71; 1.01-2.92 (0.047)
Partner has other partners	No	1.00	1.00
	Yes	2.36; 1.54-3.61 (<0.0001)	2.50; 1.60-3.90 (<0.0001)

* Not included in multivariable model

Figure 3.6 representing the map of Kenya showing the fitted individual probability of infection and the standard error map by county reveal hot spots of infection in the counties indicated by a red colour. The pattern of hot and cold spots is comparable to that obtained in 2003. In that, the hotspot counties remain the same with a few additions.

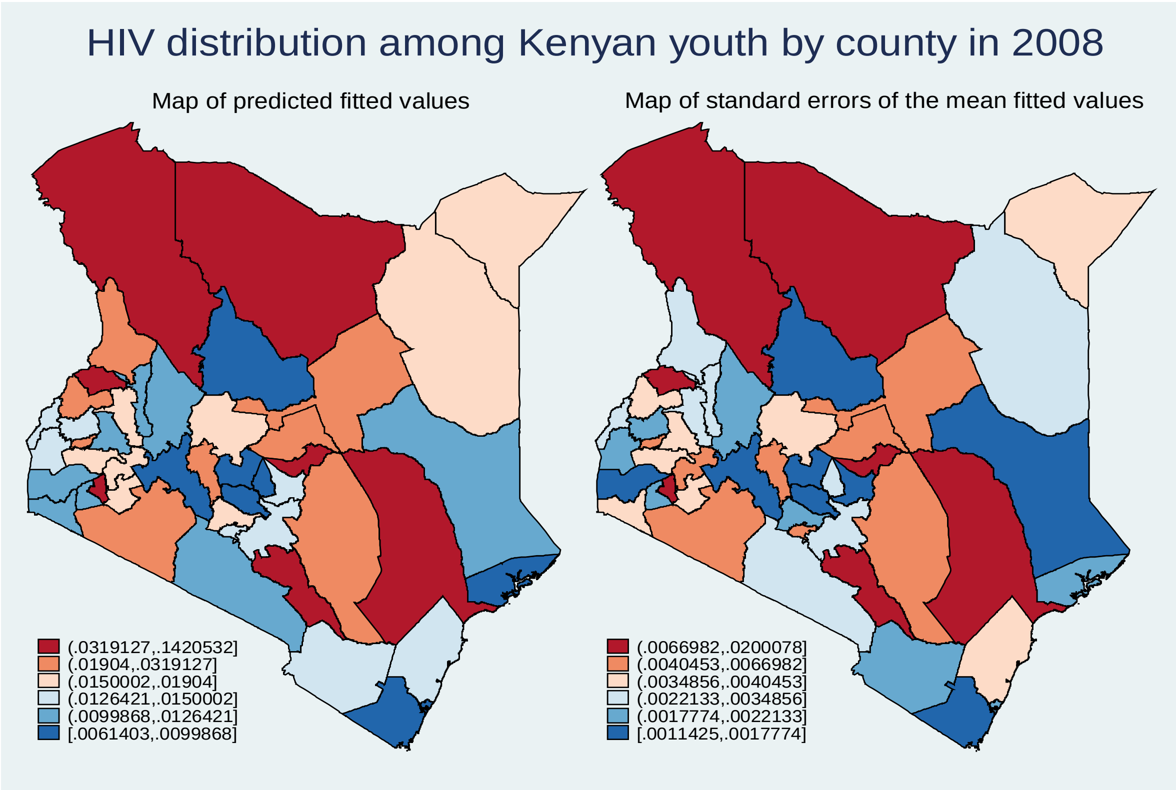


Figure 3.6 Maps showing the mean predicted fitted probability of infection and standard errors of the means by county in 2008

3.3.3 Factors associated with the spatial distribution of HIV in both years

The evidence was not sufficient to conclude that the county level distribution of HIV cases among Kenyan youth changed from 2003 to 2008 when keeping other factors constant. The year 2008 was not a significant determinant of HIV status compared to the year 2003 (spatially adjusted OR 1.03 95%CI 0.97-1.09). The results are shown in table 3.8.

The effect of socio-demographic factors such as age and gender remained the same when the data from both surveys were combined. However, the respondents' residence became a significant determinant of HIV infection (AOR 1.96 95%CI 1.36-2.81). The results from the maximum likelihood regression model with only the unstructured county random effects were comparable with only minimal differences to those from the Bayesian based model also with only unstructured county random effects. Both models were also similar to the spatial model with both unstructured and structured county random effects.

Table 3.8: Factors associated with the spatial distribution of HIV in 2003 and 2008

		Maximum likelihood based model			Bayesian-based models			
Variable	**Category**	**Multilevel logistic regression model**			**Model with only unstructured random effects**		**Model with unstructured & structured random effects**	
		AOR	**(95% CI)**	**P value**	**POR**	**(95% CI*)**	**POR**	**(95% CI*)**
Age		1.23	1.16-1.30	<0.0001	1.23	1.17-1.31	1.23	1.16-1.31
Gender	Male		***1.00***			***1.00***		***1.00***
	Female	3.33	2.26-4.92	<0.0001	3.34	2.29-4.99	3.32	2.26-4.94
Residence	Rural		***1.00***			***1.00***		***1.00***
	Urban	1.97	1.37-2.84	<0.0001	1.96	1.37-2.80	1.96	1.36-2.81
Educational level	primary		***1.00***			***1.00***		***1.00***
	no education	0.95	0.50-1.82	0.878	0.93	0.46-1.75	0.92	0.46-1.76
	secondary or higher	0.48	0.33-0.70	<0.0001	0.47	0.32-0.68	0.47	0.32-0.68
Risky sex at debut	No		***1.00***			***1.00***		***1.00***
	Yes	1.62	1.12-2.34	0.011	1.62	1.11-2.33	1.64	1.13-2.36

		Maximum likelihood based model			Bayesian-based models			
Variable	Category	Multilevel logistic regression model			Model with only unstructured random effects		Model with unstructured & structured random effects	
Unfaithful partner	No		***1.00***		***1.00***		***1.00***	
	Yes	2.11	1.53-2.90	<0.0001	2.16	1.60-2.95	2.26	1.63-3.12
No. of partners in 12m		1.02	0.96-1.08	0.572	0.98	0.85-1.06	0.98	0.85-1.06
Survey year	2003		***1.00***		***1.00***		***1.00***	
	2008	0.87	0.62-1.22	0.426	0.99	0.99-1.00	1.03	0.97-1.09
Cross-generational partner	No		***1.00***		***1.00***		***1.00***	
	Yes	1.22	0.82-1.83	0.331	1.55	1.00-2.21	1.48	0.99-2.17
Transactional Sex	No		***1.00***		***1.00***		***1.00***	
	Yes	1.76	1.02-3.01	0.041	1.75	1.00-2.91	1.80	1.02-3.05

CI* Credible interval. Bayesian model with unstructured random effects DIC=1447.960. Bayesian model with both unstructured and structured random effects DIC=1455.860

Notably, the effect of some behavioural factors such as having a partner with concurrent partners and engaging in risky sex at first sexual encounter became more pronounced in the model incorporating the structured spatial random errors. For instance, the likelihood of being positive among respondents who engaged in risky sex at first sexual encounter compared to those who did not was greater by 2% in the spatially structured model (AOR 1.64; 95% CI 1.13-2.36) compared to the unstructured model (AOR 1.62; 95%CI 1.12-2.34).

A similar observation was made among respondents who reported having a partner who practises concurrency. In the spatially adjusted model, the likelihood of being HIV positive increased by 10% among respondents who reported partners practising concurrency compared to those who did not (AOR 2.26; 95% CI 1.63-3.12) compared to (AOR 2.16; 95%CI 1.60-2.95) in the Bayesian model controlling for only unstructured random effects and other fixed factors. The evidence was insufficient to conclude that having a cross-generational partner increased the odds of being HIV positive in both the spatial and non-spatial models.

The probability of being HIV infected varied considerably by county among Kenyan youth as illustrated in figure 3.7. The results of Bayesian modelling revealed several counties with the highest posterior individual probabilities of HIV infection (indicated by a red colour) as well marked county spatial heterogeneity in the random effects on the HIV distribution.

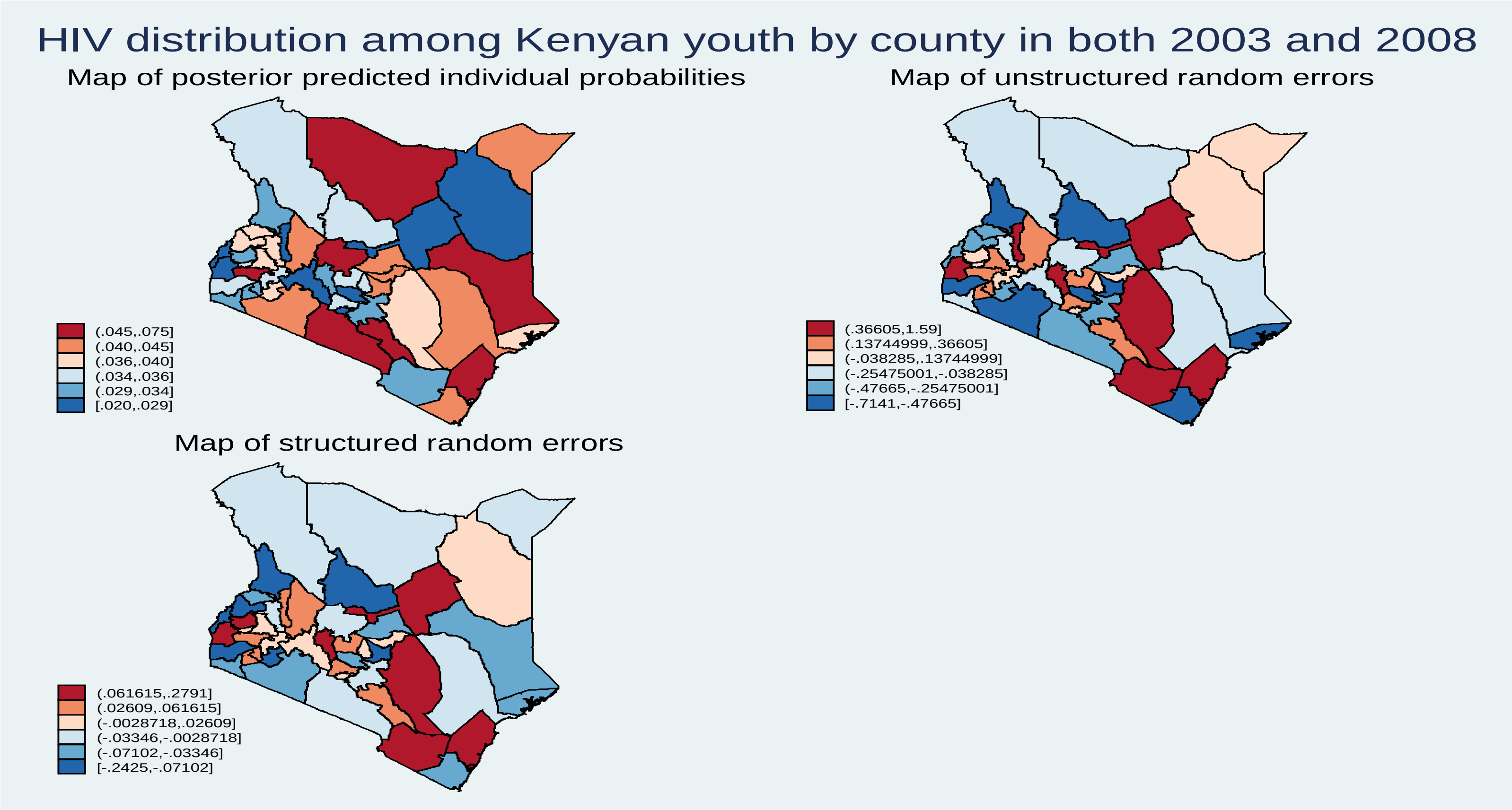


Figure 3.7 Maps showing the mean posterior individual probabilities of infection, unstructured and structured county random errors in both 2003 and 2008

3.4 Effect of sexual behaviour on HIV distribution

In the multilevel GSEM, Individual-level demographic variables such as age and sex were omitted because prior descriptive analysis showed that these variables did not change significantly between the two years hence their effect would remain the same.

Figure 3.8 shows the direct and indirect pathways from covariates on the outcome HIV status while adjusting for multilevel county random effects. In the pathway analysis, sexual risk behaviour variables significantly impacted HIV status directly as conferred by two pathways; from engaging in risky sex at sexual debut and having a partner with other concurrent sexual partners. The results are also shown in table 3.9.

These two variables both had significant positive coefficients impacting HIV status. This means that the likelihood of being HIV positive was increased similar to results obtained from classical modelling.

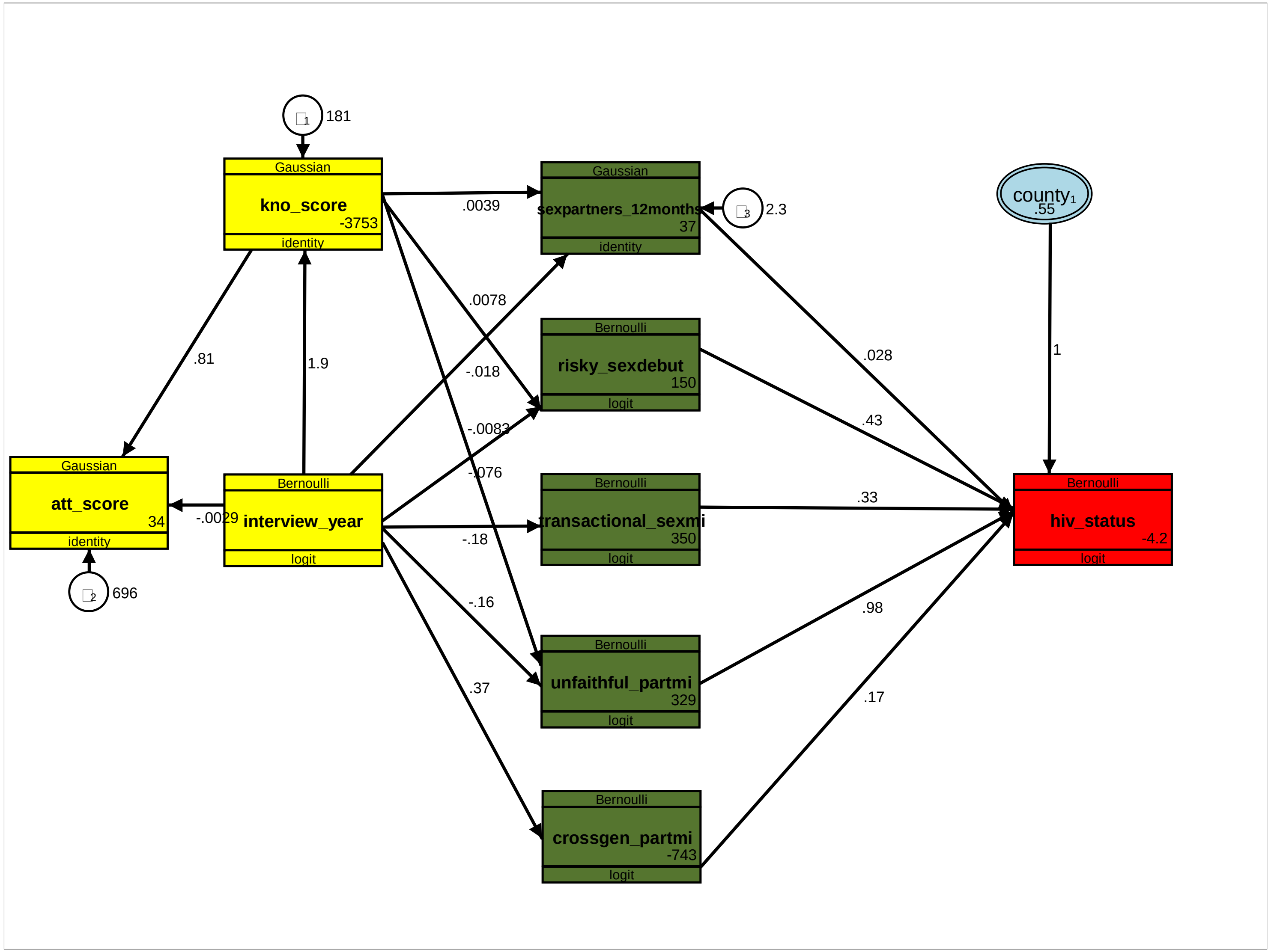


Figure 3.8 GSEM pathways results are showing coefficients for predictors of HIV status among Kenyan youth in both 2003 and 2008. Key for variable names: number of sex partners in past 12 months (**sexpartners_12months**), having risky sex at first sexual encounter (**risky_sexdebut)**, giving or receiving money/gifts for sex

Table 3.9 Factors with direct, indirect and total effects on HIV distribution among Kenyan youth in 2003 and 2008

	GSEM output showing the direct, indirect, and total effects of the determinants of HIV infection								
	Direct effects on HIV infection as shown in Fig. 3.7 GSEM diagram							Indirect effects on HIV infection	Total effects on HIV infection
	Kno score	Sexpartners _12m	Risky sexdebut	Transactional sexmi	Unfaithful _partmi	Crossgen _partmi	HIV status		
Kno_score		0.38 x10^{-2}**	0.78 x10^{-2}**		-0.83 x10^{-2}***			-0.46 x10^{-2}*	-0.13 x10^{-2}*
Att_score									
Interview_year	1.90***	-0.02*	-0.08***	-0.18***	-0.16***	0.37***		-19.69 x10^{-2}*	0.13 *
Sexpartners_12m							0.03		
Risky_sexdebut							0.43*		
Transactional_sexmi							0.33		
Crossgen_partmi							0.17		
Unfaithful_partmi							0.98***		
county							1.00		

*** P≤ 0.05, ** P≤ 0.01, *** P≤ 0.001**

Further to the direct effects of the sexual behaviour variables on HIV cases, there were also two indirect pathways that significantly impacted HIV status while controlling for county random effects. These were paths from the knowledge score and the year of the survey. Taking risky sex at sexual debut as an endogenous variable, the survey year, as well as the knowledge score, had an indirect protective effect on HIV status. Additionally, while taking having a partner with other sexual partners as an endogenous variable, the knowledge score and survey year also had an indirect protective effect on HIV infection.

Overall, the total effects from the sum of all indirect paths from comprehensive HIV knowledge score and the year of the survey both negatively impacted HIV cases. However, the greatest impact on HIV distribution was from the direct effect of having a partner who practices concurrency.

Chapter 4 Discussion, conclusion and Recommendations

This chapter contains a contextual review of the pertinent findings as well as the strengths and limitations of this research study. It also concludes the research report and provides recommendations for public health interventions and researchers synthesized from the study findings.

4.1 Discussion of findings

The primary aim of this study was to determine and compare the relationship between sexual behaviour and the prevalence of HIV among young people in Kenya the time periods, 2003 and 2008. The findings demonstrate an association between sexual risk behaviour and HIV prevalence among Kenyan youth aged 15-24 years affected by a combination of demographic, psychosocial and spatial-temporal determinants in both study years

4.1.1 Sociodemographic and psychosocial factors associated with HIV distribution

This study demonstrates that HIV distribution among Kenyan youth is associated with demographic factors such as age, gender, educational attainment and the place of residence.

The findings reveal consistent differences in HIV risk between young females and males. In both years analysed, gender was shown to be an independent predictor of HIV status as females had a higher likelihood of being HIV positive while controlling for other factors including spatial heterogeneity. This is consistent with evidence from Kenya as well as other countries in SSA (34,51,52). For instance, using nationally representative surveys in South Africa, the incidence among young females aged 15-24 years was almost five times the incidence among their male peers (51).

Apart from gender, place of residence was also a significant factor associated with HIV as youth residing in urban areas were more likely to be HIV positive compared to their rural counterparts.

Furthermore, evidence from this study also revealed that the likelihood of being HIV positive increased significantly for every year the respondents aged within the 15-24 years age group. Other studies have also shown that among young people, HIV prevalence and incidence increases with age (7,8,51).

This study also revealed that secondary or higher educational attainment was consistently protective against HIV infection in both years analysed compared to having an only primary level education. Previous surveys in Kenya have also demonstrated the lowest HIV prevalence among youth with secondary and higher educational levels as well as an association between the level of educational attainment and sexual risk behaviour among youth (7,16). Moreover, an individual's educational level may also affect their knowledge of HIV transmission and prevention.

The conceptual framework pathway analysis demonstrates that indirect determinants of sexual behaviour such as comprehensive HIV knowledge also have an impact on HIV status. Although having comprehensive knowledge on HIV transmission and prevention does not directly impact HIV status, it is associated with safer sexual behaviour. The knowledge score had a significant impact on the reduction of sexual risk behaviour such as the number of sexual partners in previous 12 months, risky sex at sexual debut and having a partner who practices concurrency.

Moreover, improving the levels of comprehensive HIV knowledge among young people over time also impacts HIV acquisition. In this analysis, the year of the survey had a bearing on knowledge levels in that knowledge levels among youth increased from 2003 to 2008. Higher levels of knowledge consequently impacted sexual behaviour causing a reduction in the proportions reporting certain risky sexual behaviour. A study on knowledge, attitudes and practices regarding HIV among young people in Lao revealed similar findings of a positive impact of higher knowledge levels on adoption of safer sexual practices (65).

4.1.2 Sexual risk behaviour associated with HIV distribution

Another significant finding is that at an individual-level, sexual risk behaviour among the youth, especially engaging in risky sex at sexual debut and having a partner who practices concurrency significantly increases the likelihood of being HIV infected while accounting for county spatial heterogeneity in Kenya.

A previous review of multiple population-based surveys conducted in SSA suggests that concurrency significantly increases the risk of HIV acquisition among partners of those who practise it (24). In this study having a partner who had other concurrent sexual partners was the greatest sexual risk determinant of HIV infection after controlling for both the direct and indirect effects of other factors.

Engaging in risky sex at first sexual encounter also had a significant impact on HIV distribution. This study investigates sexual debut as a factor that includes condom use and the timing of the first sexual encounter. The descriptive analysis reveals that the majority of Kenyan youth are sexually experienced with a very low proportion reporting condom use at sexual debut. Additionally, among HIV positive youth, the proportion who reported an early sexual debut was significantly higher than among those who were HIV negative. An early sexual debut has been linked to risky sexual behaviour and a higher risk of sexually transmitted infections including HIV (6,17).

This study did not provide sufficient evidence to conclude that the odds of having HIV increased for every additional sexual partner in the 12 months preceding the surveys while controlling for other sexual risk behaviour and demographic factors. This is consistent with other studies that have argued that concurrency presents a bigger risk of HIV transmission as opposed to having multiple sexual partners while practising serial monogamy (21).

Results from both study years indicate that while certain behavioural practices were not significantly associated with HIV status in 2003, their significance became more pronounced in 2008.

For example, engaging in transactional sex was the greatest predictor of HIV status while adjusting for spatial heterogeneity and other factors in 2008. Also, the map showing the predicted individual odds of infection in 2008 revealed hotspots in some counties surrounding Lake Victoria. This is consistent with other evidence from Kenya that suggests that the high HIV prevalence in the region results from transactional relationships (7,30,45). Moreover, certain characteristics of transactional relationships among young people particularly the differential power dynamics to negotiate safer sex in light of declining fishing opportunities have contributed to the increased odds of HIV infection as demonstrated by other studies (32,33).

In this study, we fail to observe a significant effect of having a cross-generational partner on HIV status while accounting for other sexual risk behaviour socio-demographic factors and spatial heterogeneity. This complements other studies that have failed to observe an increased risk of HIV among young people reporting cross-generational partnerships (40,41). However, this finding is not conclusive as the suggestions that the characteristics of such partnerships may be the cause of the increased HIV risk observed in some studies should be further explored.

4.1.3 Spatial variability of HIV cases among Kenyan youth

This study also reveals significant spatial variability in HIV distribution among young people in Kenya who are considered a priority population sub-group (4). A review of the available literature did not find any studies that have focused on the spatial distribution of HIV among young people in Kenya.

However, there are studies on the spatial distribution of HIV among the general population, and these studies have demonstrated similar findings (30,45). Contrastingly, the effects of age on the spatial HIV distribution among the general population is non-linear as shown in one study (45). Prevalence is shown to rise with increasing age up to a peak when it starts to decline, this study, however, shows the linear effects of age within the age group as the odds of infection significantly increase for every one year rise in age.

Consistently, the Western region of Kenya has been shown to have a disproportionate HIV burden compared to other regions of the country (7,8,30,45). The region is already prioritized in the country's new approach to HIV prevention (4). In this study, besides hotspot counties in the Western region of Kenya, other counties in Northern and Central Kenya are also identified as hotspots of infection among the youth.

4.2 Strengths and limitations of the study

A scrutiny of the classification of exposures and outcomes, methods used to deal with confounding and effect modification and the local or global relevance of findings detail the strengths and limitations of this research study.

A major strength of this study is the use of various statistical techniques to deal with confounding and effect modification. Multiple survey regression models and multilevel modelling with the inclusion of structured spatial random effects in the model were employed, and further checks for the model goodness of fit were performed. These statistical methods allow inferences to be made at the individual level considering the complex hierarchical nature of the DHS data.

Additionally, this study also investigated the indirect and total effects of covariates on HIV status using GSEM pathway analysis. The pathway analysis provided benefit beyond using only ordinary regression to determine direct effects of sexual risk behaviour on HIV distribution. Comprehensive knowledge on HIV and the temporal aspect in changing sexual behaviour and improving comprehensive HIV knowledge were identified as underlying factors that affect HIV distribution among Kenyan youth.

Another strength of the study is in the classification of the outcome and exposures to reduce measurement errors. In the original surveys, blood samples were tested using multiple tests and by different laboratories to reduce errors in classification of HIV status. HIV DNA PCR an accurate test was employed in cases of doubt. During modelling, unnecessary categorization of continuous exposure variables was avoided and where appropriate defined using existing literature. Also, sensitivity analysis of the created composite knowledge and attitude scores with similar scores created using exploratory and confirmatory factor analysis was done.

A major limitation of this study as pertains the classification of exposures and outcomes is non-response and social desirability bias. Almost 40% of the sample of young people surveyed had missing HIV test results which may indicate a refusal to get tested. This may introduce bias especially if the refusal is linked to HIV status.

There was also a significant proportion of missing values of sexual behaviour variables which may have resulted from respondents declining to answer questions probing sexual conduct. Moreover, even among those who responded, responses may have been influenced by a desire to be socially correct.

Respondents in the setting of a DHS survey are less likely to report sexual behaviour that is deemed socially inappropriate especially when family and sexual partners may be in the vicinity during an interview. This study attempts to address the nonresponse bias through multiple imputations of missing values in the variables representing sexual risk behaviour.

The study also fails to demonstrate a causal link between sexual behaviour and the HIV distribution due to the study design. A cross-sectional design implies that inferences can only be made on associations rather than causation.

Despite these limitations, these findings are significant locally in the Kenyan context as psychosocial factors and sexual behaviour are not static as was shown in this analysis. Differences may exist in individual knowledge and sexual behaviour over time and inclusion of the temporal factor in this study is useful in revealing trends in specific areas that may not be obviously apparent. Young people are also the primary focus of this analysis as they are driving new infection patterns regionally. Hence these results may be generalisable to other countries in SSA with similar HIV prevalence patterns.

4.3 Conclusion

In this research study, individual sexual risk behaviours among young people were significant determinants of HIV prevalence in Kenya in 2003 and 2008 after controlling for other relevant factors. As already demonstrated in the literature, young people continue to engage in risky sexual behaviour contributing substantially to new HIV cases. In this regard, there is still a role for behavioural intervention programs among this age group.

Specifically tailored programs, such as those encouraging young people to delay sexual debut or avoid partners practising sexual concurrency may provide substantial gains in the control of the HIV epidemic in the Kenyan context. This may be achieved through increasing comprehensive HIV knowledge levels among youth. The heterogeneity in HIV prevalence observed among counties implies that concentration of efforts on the most affected geographical areas before eventual inclusion of other regions is a reasonable public health approach.

4.4 Recommendations

Several recommendations arise from this research. These include;

Behavioural intervention programs among young people should target particular behaviour and should be geographically targeted to improve effectiveness.

Emphasis on the reduction of sexual risk behaviour among young people should be gender specific as young females are a higher risk group compared to their male counterparts.

GSEM is an important statistical analysis tool that could be employed in the analysis of complex relationships that may not be possible with classical modelling methods.

Further research on the Kenyan context with recent data upon conducting a nationally representative survey that includes HIV testing would be useful to analyse trends over a longer period.

References

1. Global AIDS Update UNAIDS. 2016. Available from: http://www.who.int/hiv/pub/arv/global-AIDS-update-2016_en.pdf [cited 2017 Jul 14].
2. The Gap Report UNAIDS. 2014. Available from: http://files.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2014/UNAIDS_Gap_report_en.pdf [cited 2016 Jul 14].
3. Cuadros DF, Abu-Raddad LJ. Spatial variability in HIV prevalence declines in several countries in sub-Saharan Africa. Health and Place. 2014;28:45–9.
4. Kenya HIV Prevention Revolution Road Map.Kenya National AIDS & STI Control Programme (NASCOP) & National AIDS Control Council (NACC). Vol. 371. 2014. Available from: http://www.gbvhivonline.com/wp-content/uploads/2014/04/HIV-prevention-roadmap-report-draft.pdf [cited 2017 Aug 8].
5. Definition of youth UN. Available from: http://www.un.org/esa/socdev/documents/youth/fact-sheets/youth-definition.pdf [cited 2016 Jul 7].
6. Idele P, Gillespie A, Porth T, Suzuki C, Mahy M, Kasedde S, et al. Epidemiology of HIV and AIDS Among Adolescents. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2014;66(SUPPL. 2):S144–53.
7. Kenya Demographic and Health Survey 2008-09. Kenya National Bureau of Statistics (KNBS); ORC Macro Health (San Francisco). 2010. Available from: http://dhsprogram.com/pubs/pdf/fr229/fr229.pdf [cited 2016 Jun 26].

8. Kenya Demographic and Health Survey 2003 Central Bureau of Statistics, ICF Macro. 2004. Available from: http://www.dhsprogram.com/pubs/pdf/FR151/FR151.pdf [cited 2016 Jun 26].

9. Kenya AIDS Indicator Survey 2012 National AIDS and STI Control Programme (NASCOP) K 2014. Available from: http://nacc.or.ke/wp-content/uploads/2015/10/KAIS-2012.pdf [cited 2016 Aug 17].

10. AIDSinfo UNAIDS 2015. Available from: http://aidsinfo.unaids.org/ [cited 2016 Sep 7].

11. Kharsany ABM, Karim QA. HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities. The open AIDS journal. 2016;10:34–48.

12. Marston C, King E. Factors that shape young people's sexual behaviour: a systematic review. The Lancet. 2006;368(9547):1581–6.

13. Wellings K, Collumbien M, Slaymaker E, Singh S, Hodges Z, Patel D, et al. Sexual behaviour in context: a global perspective. Vol. 368, Lancet. 2006. p. 1706–28.

14. Bearinger LH, Sieving RE, Ferguson J, Sharma V. Global perspectives on the sexual and reproductive health of adolescents: patterns, prevention, and potential. Lancet. 2007;369(9568):1220–31.

15. Opportunity in Crisis-Preventing HIV from early adolescence to young adulthood. UNICEF. 2011. Available from: https://www.unicef.org/publications/files/Opportunity_in_Crisis-Report_EN_052711.pdf [cited 2016 Aug 7].

16. Demographic K. Health Survey 2014: key indicators. Kenya National Bureau of Statistics (KNBS) and ICF Macro. 2015. Available from: https://dhsprogram.com/pubs/pdf/fr308/fr308.pdf [cited 2016 Sep 7].
17. Stöckl H, Kalra N, Jacobi J, Watts C. Is Early Sexual Debut a Risk Factor for HIV Infection Among Women in Sub-Saharan Africa? A Systematic Review. Vol. 69, American Journal of Reproductive Immunology. 2013. p. 27–40.
18. Buttmann N, Nielsen A, Munk C, Frederiksen K, Liaw K-L, Kjaer SK. Young age at first intercourse and subsequent risk-taking behaviour: An epidemiological study of more than 20,000 Danish men from the general population. Scandinavian Journal of Public Health. 2014 Aug 1;42(6):511–7.
19. Epstein M, Bailey JA, Manhart LE, Hill KG, Hawkins JD, Haggerty KP, et al. Understanding the link between early sexual initiation and later sexually transmitted infection: Test and replication in two longitudinal studies. Journal of Adolescent Health. 2014 Apr;54(4):435–41.
20. Epstein H, Morris M. Concurrent partnerships and HIV: An inconvenient truth. Vol. 14, Journal of the International AIDS Society 2011,14:13.
21. Morris M, Kretzschmar M. A microsimulation study of the effect of concurrent partnerships on the spread of HIV in Uganda. Mathematical Population Studies. 2000;8(2):109–33.
22. UNAIDS Terminology Guidelines UNAIDS. 2015. Available from: http://www.unaids.org/sites/default/files/media_asset/JC2118_terminology-guidelines_en_1.pdf [cited 2017 Jul 4].

23. Hotton AL, French AL, Hosek SG, Kendrick SR, Lemos D, Brothers J, et al. Relationship Dynamics and Sexual Risk Reduction Strategies Among Heterosexual Young Adults: A Qualitative Study of Sexually Transmitted Infection Clinic Attendees at an Urban Chicago Health Center. AIDS Patient Care and STDs. 2015;29(12):668–74.
24. Mishra V, Assche SB. 2009 Concurrent Sexual Partnerships and HIV Infection : Evidence from National Population-Based Surveys. DHS Working Papers No. 62. Calverton, Maryland: Macro International Inc.
25. Stoebenau K, Heise L, Wamoyi J, Bobrova N. Revisiting the understanding of ‘transactional sex’ in sub-Saharan Africa: A review and synthesis of the literature. Social Science and Medicine. 2016 Nov;168:186–97.
26. Wallmyr G, Welin C, Remes P, Mbwambo J, Zhuwau T, Caraël M, et al. Young People, Pornography, And Sexuality: Sources And Attitudes. The Journal of School Nursing. 2006 Oct 15;22(5):290–5.
27. Dunkle KL, Jewkes R, Nduna M, Jama N, Levin J, Sikweyiya Y, et al. Transactional sex with casual and main partners among young South African men in the rural Eastern Cape: Prevalence, predictors, and associations with gender-based violence. Social Science & Medicine. 2007;65:1235–48.
28. Kamndaya M, Vearey J, Thomas L, Kabiru CW, Kazembe LN. The role of material deprivation and consumerism in the decisions to engage in transactional sex among young people in the urban slums of Blantyre, Malawi. Global Public Health. 2016;11(3):295–308.

29. Kamndaya M, Kazembe LN, Vearey J, Kabiru CW, Thomas L. Material deprivation and unemployment affect coercive sex among young people in the urban slums of Blantyre, Malawi: A multi-level approach. Health and Place. 2015;33:90–100.

30. Hoshi T, Fuji Y, Nzou SM, Tanigawa C, Kiche I, Mwau M, et al. Spatial distributions of HIV infection in an endemic area of Western Kenya: Guiding information for localized HIV control and prevention. PLoS ONE. 2016;11(2).

31. Ondondo RO, Waithera Z, Mpoke S, Kiptoo M, Bukusi EA. Prevalence and Incidence of HIV Infection among Fishermen along Lake Victoria Beaches in Kisumu County, Kenya. World Journal of AIDS. 2014;4(June):219–31.

32. Kwena ZA, Bukusi E, Omondi E, Ng'ayo M, Holmes KK. Transactional sex in the fishing communities along Lake Victoria, Kenya: a catalyst for the spread of HIV. African Journal of AIDS Research. 2012 Mar;11(1):9–15.

33. Fiorella KJ, Camlin CS, Salmen CR, Omondi R, Hickey MD, Omollo DO, et al. Transactional Fish-for-Sex Relationships Amid Declining Fish Access in Kenya. World Development. 2015 Oct;74:323–32.

34. Mdodo R, Gust D, Otieno FO, McLellan-Lemal E, Chen RT, LeBaron C, et al. Investigation of HIV Incidence Rates in a High-Risk, High-Prevalence Kenyan Population. Journal of the International Association of Providers of AIDS Care (JIAPAC). 2016;15(1):42–50.

35. Harrison A, Colvin CJ, Kuo C, Swartz A, Lurie M. Sustained High HIV Incidence in Young Women in Southern Africa: Social, Behavioral and Structural Factors and Emerging Intervention Approaches. Current HIV/AIDS Reports. 2015;12(2):207–15.
36. Longfield K, Glick A, Waithaka M, Berman J. Relationships Between Older Men and Younger Women: Implications for STIs/HIV in Kenya. Studies in Family Planning. 2004 Jun 1;35(2):125–34.
37. Akullian A, Bershteyn A, Klein D, Vandormael A, Bärnighausen T, Tanser F. Sexual partnership age-pairings and risk of HIV acquisition in rural South Africa. AIDS. 2017 Jul 31;31.(12) 1755-1764.
38. Pettifor AE, Rees H V, Kleinschmidt I, Steffenson AE, MacPhail C, Hlongwa-Madikizela L, et al. Young people's sexual health in South Africa: HIV prevalence and sexual behaviors from a nationally representative household survey. AIDS. 2005 Sep 23;19(14):1525–34.
39. Kaiser R, Bunnell R, Hightower A, Kim AA, Cherutich P, Mwangi M, et al. Factors associated with HIV infection in married or cohabitating couples in Kenya: Results from a nationally representative study. PLoS ONE. 2011 Mar 15;6(3):e17842.
40. Harling G, Newell M-L, Tanser F, Kawachi I, Subramanian SV, Bärnighausen T. Do Age-Disparate Relationships Drive HIV Incidence in Young Women? Evidence from a Population Cohort in Rural KwaZulu-Natal, South Africa. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2014 Aug 1;66(4):443–51.

41. Balkus JE, Nair G, Montgomery ET, Mishra A, Palanee-Phillips T, Ramjee G, et al. Age-Disparate Partnerships and Risk of HIV-1 Acquisition Among South African Women Participating in the VOICE Trial. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2015 Oct;70(2):212–7.
42. Luke N. Confronting the 'sugar daddy' stereotype: age and economic asymmetries and risky sexual behavior in urban Kenya. International family planning perspectives. 2005;31(1):6–14.
43. Ritchwood TD, Hughes JP, Jennings L, MacPhail C, Williamson B, Selin A, et al. Characteristics of Age-Discordant Partnerships Associated With HIV Risk Among Young South African Women (HPTN 068). Journal of acquired immune deficiency syndromes (1999). 2016 Aug 1;72(4):423–9.
44. Morrison-Beedy D, Xia Y, Passmore D. Sexual risk factors for partner age discordance in adolescent girls and their male partners. Journal of Clinical Nursing. 2013 Dec 1;22(23-24):3289–99.
45. Ngesa O, Mwambi H, Achia T. Bayesian spatial semi-parametric modeling of HIV variation in Kenya. PLoS ONE. 2014 Jul 25;9(7):e103299.
46. Xu Y, Yang G, Liu H, Li X, Song L, Li Y, et al. Epidemiologic features of AIDS in China, 2004-2013. Clinical Infectious Diseases. 2015 Jan 1;60(1):167–9.
47. Zulu LC, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010. BMC Infectious Diseases. 2014;14(1):285.

48. Westercamp N, Moses S, Agot K, Ndinya-Achola JO, Parker C, Amolloh KO, et al. Spatial distribution and cluster analysis of sexual risk behaviors reported by young men in Kisumu, Kenya. International Journal of Health Geographics. 2010;9(1):24.
49. Kim AA, Parekh BS, Umuro M, Galgalo T, Bunnell R, Makokha E, et al. Identifying risk factors for recent HIV infection in Kenya using a recent infection testing algorithm: Results from a nationally representative population-based survey. Dezzutti CS, editor. PLoS ONE. 2016 May 19;11(5):e0155498.
50. Aral SO, Torrone E, Bernstein K. Geographical targeting to improve progression through the sexually transmitted infection/HIV treatment continua in different populations. Current Opinion in HIV and AIDS. 2015;10(6):477–82.
51. Shisana O, Rhele T, Simbayi LC, Zuma K, Jooste S, Zungu N, et al. South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. Available from: http://www.hsrc.ac.za/en/research-data/view/6871 [cited 2017 Aug 21].
52. Ghys E, Gouws RL, Garcia-Calleja J, Barrerre B C, D Serrano, Velasquez. Trends in HIV prevalence and sexual behaviour among young people aged 15-24 years in countries most affected by HIV. Sexually transmitted infections. 2010 Dec;86 Suppl 2:ii72-ii83.
53. Santelli JS, Edelstein ZR, Mathur S, Wei Y, Zhang W, Orr MG, et al. Behavioral, biological, and demographic risk and protective factors for new HIV infections among youth in Rakai, Uganda. Journal of acquired immune deficiency syndromes (1999). 2013;63(3):393–400.

54. Young people, HIV and AIDS. AVERT. 2017. Available from: https://www.avert.org/node/389/pdf [cited 2017 Aug 21].

55. Shisana O, Risher K, Celentano DD, Zungu N, Rehle T, Ngcaweni B, et al. Does marital status matter in an HIV hyperendemic country? Findings from the 2012 South African National HIV Prevalence, Incidence and Behaviour Survey. AIDS care. 2016;28(2):234-41.

56. Fox AM. Marital concurrency and HIV risk in 16 African countries. AIDS and Behavior. 2014 Apr 8;18(4):791–800.

57. World Population Monitoring. United Nations. 2012. Available from: http://www.un.org/en/development/desa/population/publications/pdf/fertility/12_66976_adolescents_and_youth.pdf [cited 2017 Aug 21].

58. Akwara PA, Madise NJ, Hinde A. Perception of risk of HIV/AIDS and sexual behaviour in Kenya. Journal of Biosocial Science. 2003 Jul;35(3):385-411.

59. Bernardi L. Determinants of individual AIDS risk perception: knowledge, behavioural control and social influence. African Journal of AIDS Research. 2002;1(2):111–24.

60. Anderson KG, Beutel AM, Maughan-Brown B. HIV risk perceptions and first sexual intercourse among youth in Cape Town South Africa. International family planning perspectives. 2007;33(3):98–105.

61. Beutel AM, Anderson KG. Change over time in the Human Immunodeficiency Virus risk perceptions of youth. Journal of Social Sciences. 2013 Jan 1;9(1):1–10.

62. Tenkorang EY, Maticka-Tyndale E. Assessing young people's perceptions of HIV risks in Nyanza, Kenya: Are school and community level factors relevant? Social Science and Medicine. 2014;116:93–101.

63. Tenkorang EY, Rajulton F, Maticka-Tyndale E. Perceived risks of HIV/AIDS and first sexual intercourse among youth in Cape Town, South Africa. AIDS and Behavior. 2009 Apr 10;13(2):234–45.

64. Asaolu IO, Gunn JK, Center KE, Koss MP, Iwelunmor JI, Ehiri JE. Predictors of HIV testing among youth in sub-Saharan Africa: A cross-sectional study. PLoS ONE. 2016;11(10):e0164052.

65. Thanavanh B, Harun-Or-Rashid M, Kasuya H, Sakamoto J. Knowledge, attitudes and practices regarding HIV/AIDS among male high school students in Lao People's Democratic Republic. Journal of the International AIDS Society. 2013 Mar 11;16(1):17387.

66. Shokoohi M, Karamouzian M, Mirzazadeh A, Haghdoost A, Rafierad A-A, Sedaghat A, et al. HIV Knowledge, Attitudes, and Practices of Young People in Iran: Findings of a National Population-Based Survey in 2013. Keyhani S, editor. PLoS ONE. 2016 Sep 14;11(9):e0161849.

67. Ewomazino M. Youths and Risky Sexual Behaviour : A Kap Study on Hiv / Aids amongst University of Ibadan Student Youths and Risky Sexual Behaviour : A Kap Study on HIV/AIDS amongst University of Ibadan Student. European Journal of Social Sciences. 2016;49(October 2015):140–52.

68. Santelli JS, Edelstein ZR, Wei Y, Mathur S, Song X, Schuyler A, et al. Trends in HIV acquisition, risk factors and prevention policies among youth in Uganda,

1999-2011. AIDS (London, England). 2015 Jan;29(2):211–9.

69. Kembo J. Changes in sexual behaviour and practice and HIV prevalence indicators among young people aged 15–24 years in Zambia: An in-depth analysis of the 2001–2002 and 2007 Zambia Demographic and Health Surveys. Sahara-J: Journal of Social Aspects of HIV/AIDS. 2013;10(3-4):150–62.

70. Hallett TB. Declines in HIV prevalence can be associated with changing sexual behaviour in Uganda, urban Kenya, Zimbabwe, and urban Haiti. Sexually Transmitted Infections. 2006 Apr;82(suppl_1):i1–8.

71. Tanser F, de Oliveira T, Maheu-Giroux M, Bärnighausen T. Concentrated HIV subepidemics in generalized epidemic settings. Current Opinion in HIV and AIDS. 2014 Mar;9(2):115–25.

72. Constitution of Kenya. Kenya Law Reform Commission (KLRC). Available from: http://www.klrc.go.ke/index.php/constitution-of-kenya/138-chapter-eleven-devolved-government [cited 2017 Jul 4].

73. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. International Journal of Epidemiology. 1999;28(2):319–26.

74. Hijmans R. Global Administrative Areas-Boundaries without limits. 2009. Available from: http://www.gadm.org/ [cited 2017 Aug 22].

75. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS–A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing. 2000;10(4):325–37.

76. R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.r-project.org/

77. Using DataSets for Analysis-The DHS Program MEASURE DHS. Available from: http://dhsprogram.com/data/Using-DataSets-for-Analysis.cfm#CP_JUMP_14042 [cited 2017 Sep 1].

78. Safety of injections. WHO. Vol. 37, 2000; p. 345–7. Available from: http://who.int/injection_safety/about/resources/en/QuestionAndAnswersInjectionSafety.pdf [cited 2017 Aug 22].

79. CDC. Basics | Blood Safety | CDC 2013. Available from: https://www.cdc.gov/bloodsafety/basics.html [cited 2017 Aug 22].

80. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009 Jun 29;338:b2393.

81. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? International journal of methods in psychiatric research. 2011 Mar;20(1):40–9.

82. Weesie J. suest-Seemingly unrelated estimation. Available from: http://www.stata.com/manuals13/rsuest.pdf [cited 2017 Aug 29].

83. Pisati M. SPMAP: Stata module to visualize spatial data. Available from: https://ideas.repec.org/c/boc/bocode/s456812.html [cited 2017 Sep 4].

84. Spiegelhalter D, Thomas A, Best N, Way R. WinBUGS User Manual. Components. 2003;2(January):1–60.

85. Miller HJ. Tobler’s first law and spatial analysis. Annals of the Association of American Geographers. 2004 Jun 1;94(2):284–9.

86. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002 Oct 1;64(4):583–639.

87. Rabe-Hesketh S, Skrondal A, Pickles A. Generalized Multilevel Structural Equation Modeling. Psychometrika. 2004;69(2):167–90.

Appendix A

A.1 Plagiarism declaration form

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I ___Brenda Anyango Oulo____ (Student number: _1508068_) am a student registered for the degree of __Master of science Infectious Disease Epidemiology_ in the academic year _2017_.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: ______ ___Date: ___28/02/18________________

A.2 Stata codes used in sample weighting and survey setting the data

```
**survey setting the data taking sample weights into account*

egen strata=group(region residence)

gen sampwt=sample_weight/1000000

gen psu=clustno

svyset psu [pw=sampwt],strata(strata)
```

A.3 Correlation of scores obtained using structural equation modelling (SEM) and manual summation method

A.3.1 Knowledge factor scores using SEM for 2008

```
sem (Knowledge -> kno_MTC_pregnancy, ) (Knowledge -> kno_MTC_delivery, )
(Knowledge -> kno_MTC_breastfeeding, ) (Knowledge -> kno_PMTCT, )
(Knowledge -> kno_abstain, ) (Knowledge -> kno_conduse, ) (Knowledge ->
kno_limit_onepartner, ) (Knowledge -> kno_limitpartners, ) (Knowledge ->
kno_avoidsexprostitutes, ) (Knowledge -> kno_avoidperson_manypartners, )
(Knowledge -> kno_avoidgaysex, ) (Knowledge -> kno_avoidsex_drugusers, )
(Knowledge -> kno_avoidsharing_sharps, ) (Knowledge -> rkno_avoidmosqbites, )
(Knowledge -> rkno_noways, ) (Knowledge -> rkno_noknowledge, ) (Knowledge ->
rkno_avoidkissing, ) (Knowledge -> rkno_aidswitchcraft, ) (Knowledge ->
rkno_avoidsharefood, ), method(mlmv) latent(Knowledge ) nocapslatent
estat gof, stats(all) //acceptable model fit using CFI 0.954 , TLI 0.934, RMSEA 0.048
estat mindices
predict kno_semscore, latent(Knowledge)
```

```
. corr  kno_semscore kno_score
(obs=2857)

             | kno_se~e kno_sc~e
-------------+------------------
kno_semscore |   1.0000
   kno_score |   0.8279   1.0000
```

A.3.2 Knowledge factor scores using SEM for 2003

sem (Knowledge -> kno_MTC_delivery,) (Knowledge -> kno_MTC_breastfeeding,) (Knowledge -> kno_PMTCT,) (Knowledge -> kno_abstain,) (Knowledge -> kno_conduse,) (Knowledge -> kno_limit_onepartner,) (Knowledge -> kno_limitpartners,) (Knowledge -> kno_avoidsexprostitutes,) (Knowledge -> kno_avoidperson_manypartners,) (Knowledge -> kno_avoidgaysex,) (Knowledge -> kno_avoidsex_drugusers,) (Knowledge -> kno_avoidsharing_sharps,) (Knowledge -> rkno_avoidmosqbites,) (Knowledge -> rkno_noways,) (Knowledge -> rkno_avoidkissing,) (Knowledge -> rkno_avoidsharefood,), latent(Knowledge) nocapslatent

```
. corr  kno_semscore kno_score
(obs=2650)

             | kno_se~e kno_sc~e
-------------+------------------
kno_semscore |   1.0000
   kno_score |   0.6842   1.0000
```

A.4 Questions used to create composite scores

A.4.1 Comprehensive HIV knowledge score

1. Have you ever heard of an illness called AIDS?

2. Is there anything a person can do to avoid getting AIDS or the virus that causes AIDS?

3. Can a person avoid AIDS by abstaining from sex?

4. Can a person avoid AIDS by using condoms?

5. Can a person avoid AIDS by limiting sex to one partner?

6. Can a person avoid AIDS by faithful to one partner?

7. Can a person avoid AIDS by limiting the number of sex partners?

8. Can a person avoid AIDS by avoiding sex with prostitutes?

9. Can a person avoid AIDS by avoiding sex with persons who have many partners?

10. Can a person avoid AIDS by avoiding sex with homosexuals?

11. Can a person avoid AIDS by avoiding sex with drug users?

12. Can a person avoid AIDS by avoiding blood transfusions*

13. Can a person avoid AIDS by avoiding injections *

12. Can a person avoid AIDS by avoiding sharing razors/blades?

13. Can a person avoid AIDS by avoiding kissing?

14 Can a person avoid AIDS by avoiding mosquito bites?

15. Can a person avoid AIDS by seeking protection from traditional healer?

16. Can a person get the AIDS virus by sharing utensils with a person who has AIDS?

17. Do you know any other ways of avoiding AIDS

18. Can the virus that causes AIDS be transmitted from a mother to her child during pregnancy?

19. Can the virus that causes AIDS be transmitted from a mother to her child during delivery?

20. Can the virus that causes AIDS be transmitted from a mother to her child by breastfeeding?

21. Can a mother who is infected with the AIDS virus reduce the risk of giving the virus to the baby by taking certain drugs during pregnancy?

22. Is it possible for a healthy-looking person to have the AIDS virus?

**** Questions 12 and 13 omitted because they were not specific**

A.4.2 Attitude towards people living with HIV

1. Would you buy fresh vegetables from a vendor who has the AIDS virus?
2. If a member of your family got infected with the virus that causes AIDS, would you want it to remain a secret or not?
3. If a relative of yours became sick with the virus that causes AIDS, would you be willing to care for her or him in your own household?
4. If a female teacher has the AIDS virus, should she be allowed to continue teaching in school?

A.5 Codes and output when performing the seemingly unrelated estimation (suest)

suest B A, svy

test [B_hiv_status=A_hiv_status], common

Adjusted Wald test

(1) [B_hiv_status]age - [A_hiv_status]age = 0

(2) [B_hiv_status]1b.gender - [A_hiv_status]1b.gender = 0

(3) [B_hiv_status]2.gender - [A_hiv_status]2.gender = 0

(4) [B_hiv_status]0b.transactional_sexmi - [A_hiv_status]0b.transactional_sexmi = 0

(5) [B_hiv_status]1.transactional_sexmi - [A_hiv_status]1.transactional_sexmi = 0

(6) [B_hiv_status]0b.risky_sexdebut - [A_hiv_status]0b.risky_sexdebut = 0

(7) [B_hiv_status]1.risky_sexdebut - [A_hiv_status]1.risky_sexdebut = 0

(8) [B_hiv_status]sexpartners_12months - [A_hiv_status]sexpartners_12months = 0

(9) [B_hiv_status]0b.unfaithful_partmi - [A_hiv_status]0b.unfaithful_partmi = 0

(10) [B_hiv_status]1.unfaithful_partmi - [A_hiv_status]1.unfaithful_partmi = 0

(11) [B_hiv_status]0b.crossgen_partmi - [A_hiv_status]0b.crossgen_partmi = 0

(12) [B_hiv_status]1.crossgen_partmi - [A_hiv_status]1.crossgen_partmi = 0

(13) [B_hiv_status]1b.residence1 - [A_hiv_status]1b.residence1 = 0

(14) [B_hiv_status]2.residence1 - [A_hiv_status]2.residence1 = 0

(15) [B_hiv_status]1.educ_cat - [A_hiv_status]1.educ_cat = 0

(16) [B_hiv_status]2b.educ_cat - [A_hiv_status]2b.educ_cat = 0

(17) [B_hiv_status]3.educ_cat - [A_hiv_status]3.educ_cat = 0

(18) [B_hiv_status]2003b.interview_year - [A_hiv_status]2003b.interview_year = 0

(19) [B_hiv_status]2008.interview_year - [A_hiv_status]2008.interview_year = 0

F(11, 679) = 0.22

Prob > F = 0.9962

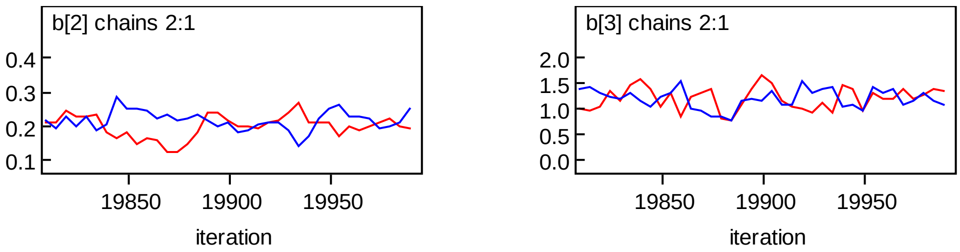
A.6 The Bayesian model

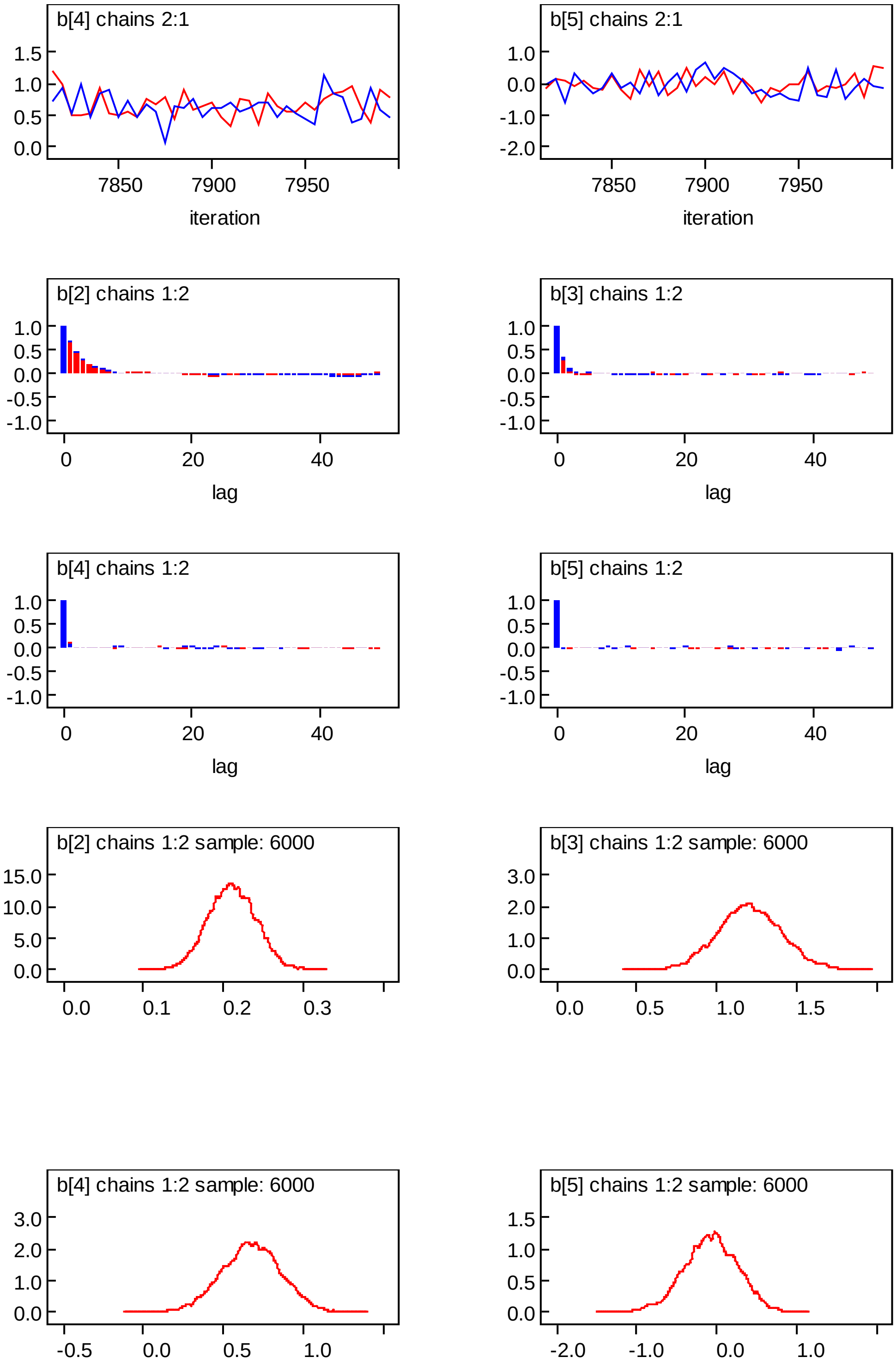
```
model{
                for (i in 1:N){
        hiv_status[i]~dbern(p[i])  # HIV main outcome variable

        logit(p[i])<-
b[1]+b[2]*age[i]+b[3]*gender[i]+b[4]*residence1[i]+b[5]*equals(educ_cat[i],1)+b[6]*equals(ed
uc_cat[i],3)+b[7]*risky_sexdebut[i]+b[8]*unfaithful_partmi[i]+b[9]*sexpartners_12months[i]+b[
10]*interview_year[i]+b[11]*crossgen_partmi[i]+b[12]*transactional_sexmi[i]+re[county[i]]+sp[
county[i]]
                                                }


for (k in 1:48) {
# Exchangeable prior on unstructured random effects
re[k] ~ dnorm(0, tau.h)
         }

# CAR prior distribution for spatial random effects:
sp[1:48] ~ car.normal(adj[], weights[], num[], tau.b)
for (k in 1:sumNumNeigh)
                              {
weights[k] <- 1
                             }


# All priors and hyperpriors:

#Priors for beta coefficients
for (i in 1:12) {
b[i]~dnorm(0,0.001)
                              }

#Priors for the unstructured and structured random effects
tau.b ~ dgamma(0.5, 0.0005)
sigma.b <- sqrt(1/tau.b)

tau.h ~ dgamma(0.5, 0.0005)
sigma.h <- sqrt(1/tau.h)


Inits chain 1 #Initial Values List 1
list(
b = c(0,0,0,0,0,0,0,0,0,0,0,0),
re =
c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
, sp = c(
0.0,0.0,0.0,0.0,0.0,
```

```
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0),
tau.b = 0.1,
tau.h = 0.1)
Inits chain 2#Initial Values List 2
list(
b = c(0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1),
re =
c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
, sp = c(
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0,0.0,0.0,
0.0,0.0,0.0),
tau.b = 1,
tau.h = 1)
```

A.7 Checking Bayesian model convergence

A sample of trace, autocorrelation and kernel density plots of select parameters




b[4] chains 2:1
1.5
1.0
0.5
0.0
7850
7900
7950
iteration
b[5] chains 2:1
1.0
0.0
-1.0
-2.0
7850
7900
7950
iteration
b[2] chains 1:2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag
b[3] chains 1:2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag
b[4] chains 1:2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag
b[5] chains 1:2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag
b[2] chains 1:2 sample: 6000
15.0
10.0
5.0
0.0
0.0
0.1
0.2
0.3
b[3] chains 1:2 sample: 6000
3.0
2.0
1.0
0.0
0.0
0.5
1.0
1.5
b[4] chains 1:2 sample: 6000
3.0
2.0
1.0
0.0
-0.5
0.0
0.5
1.0
b[5] chains 1:2 sample: 6000
1.5
1.0
0.5
0.0
-2.0
-1.0
0.0
1.0

Appendix B

B.1 Stata codes used in calculation of indirect effect pathways

The model;

gsem (kno_score -> att_score, family(gaussian) link(identity)) (kno_score -> sexpartners_12months, family(gaussian) link(identity)) (kno_score -> unfaithful_partmi, family(bernoulli) link(logit)) (kno_score -> risky_sexdebut, family(bernoulli) link(logit)) (sexpartners_12months -> hiv_status, family(bernoulli) link(logit)) (transactional_sexmi -> hiv_status, family(bernoulli) link(logit)) (unfaithful_partmi -> hiv_status, family(bernoulli) link(logit)) (crossgen_partmi -> hiv_status, family(bernoulli) link(logit)) (risky_sexdebut -> hiv_status, family(bernoulli) link(logit)) (interview_year -> kno_score, family(gaussian) link(identity)) (interview_year -> att_score, family(gaussian) link(identity)) (interview_year -> sexpartners_12months, family(gaussian) link(identity)) (interview_year -> transactional_sexmi, family(bernoulli) link(logit)) (interview_year -> unfaithful_partmi, family(bernoulli) link(logit)) (interview_year -> crossgen_partmi, family(bernoulli) link(logit)) (interview_year -> risky_sexdebut, family(bernoulli) link(logit)) (M1[county] -> hiv_status, family(bernoulli) link(logit)), covstruct(_lexogenous, diagonal) latent(M1) nocapslatent

******* separate indirect pathways from the knowledge score******

gsem, coeflegend

nlcom _b[risky_sexdebut:kno_score]*_b[hiv_status:risky_sexdebut]

nlcom _b[unfaithful_partmi:kno_score]*_b[hiv_status:unfaithful_partmi]

nlcom _b[sexpartners_12months:kno_score]*_b[hiv_status:sexpartners_12months]

*****separate indirect effects from the year of survey*****

nlcom _b[kno_score:interview_year]*

_b[unfaithful_partmi:kno_score]*_b[hiv_status:unfaithful_partmi] nlcom _b[kno_score:interview_year]*

_b[risky_sexdebut:kno_score]*_b[hiv_status:risky_sexdebut]

nlcom _b[kno_score:interview_year]*

_b[sexpartners_12months:kno_score]*_b[hiv_status:sexpartners_12months]

nlcom

_b[sexpartners_12months:interview_year]*_b[hiv_status:sexpartners_12months]

nlcom _b[risky_sexdebut:interview_year]*_b[hiv_status:risky_sexdebut]

nlcom _b[transactional_sexmi:interview_year]*_b[hiv_status:transactional_sexmi]

nlcom _b[unfaithful_partmi:interview_year]*_b[hiv_status:unfaithful_partmi]

nlcom _b[crossgen_partmi:interview_year]*_b[hiv_status:crossgen_partmi]

B.2 Ethics approval certificate

R14/49 Dr Brenda Oulo

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M1611100

NAME: (Principal Investigator)	Dr Brenda Oulo
DEPARTMENT:	School of Public Health
PROJECT TITLE:	Effect of Sexual Behaviour on the Spatial Distribution of HIV among Kenyan Youth in 2003 and 2008
DATE CONSIDERED:	25/11/2016
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr E. Musenge and Mrs L. Chimoyi-Otwombe
APPROVED BY:	Professor P Cleaton-Jones,Chairperson, HREC (Medical)
DATE OF APPROVAL:	30/11/2016

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in November and will therefore be due in the month of November each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

B.3 WinBUGS map showing structured random errors