DETERMINANTS OF HIV PREVALENCE IN KENYA

Research Report in Mathematics, 2018

AMOS EUGENE KIPTUM

August 2018

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry
DETERMINANTS OF HIV PREVALENCE IN KENYA

Research Report in Mathematics, 2018

AMOS EUGENE KIPTUM
School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master of Science Project
Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Prepared for The Director
Graduate School
University of Nairobi

Monitored by School of Mathematics
Abstract

Thirty four years ago, the first case of Human Immunodeficiency Virus was reported in Kenya. Since then, the virus has caused many deaths. It continues to be a threat to our economy considering its effects on the active members of the society. There exists wide discrepancies in HIV rates across the country, and to prioritize interventions in the fight against HIV/AIDS, it is important to build sound models that help to identify determinants of HIV prevalence. This study was conducted with the aim of finding out the fueling factors of prevalence of HIV/AIDS in Kenya. Data from The Kenya Aids Indicator Survey, 2012 was used. The data encompass male and female individuals of age groups of 15-64 years. Binary logistic regression model was used and socioeconomic and demographic factors that could have an impact on the prevalence of the disease causing virus were included in the analysis. Age and marital status had a significant effect on HIV prevalence among female and male respondent’s. The odds of being seropositive for HIV are higher in the age groups 15-24 and among married and divorced/widowed/separated individuals. Unexpectedly poverty was found to have an insignificant impact on the prevalence of HIV/AIDS among male respondents. Unlike for male respondents, wealth status was revealed to have a significant impact on the prevalence of HIV/AIDS for female individuals. Similarly, significant effects were found on the respondents, level of education, place of residence and gender. Odds of being HIV positive are significantly higher among urban dwellers and the female gender.
Declaration and Approval

I the undersigned declare that this project report is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

________________________  ________________________
Signature                Date

AMOS EUGENE KIPTUM
Reg No. 156/81947/2015

In my capacity as a supervisor of the candidate, I certify that this report has my approval for submission.

________________________  ________________________
Signature                Date

DR. IDAH OROWE
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: orowe@uonbi.ac.ke
Dedication

The project is dedicated to my girls, Julie Awuor and Amira Kiptum.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Declaration and Approval</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>vii</td>
</tr>
<tr>
<td>Figures and Tables</td>
<td>x</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xi</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>1</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>2</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>2</td>
</tr>
<tr>
<td>1.2 HIV prevalence in Kenya</td>
<td>2</td>
</tr>
<tr>
<td>1.3 HIV prevalence and gender</td>
<td>2</td>
</tr>
<tr>
<td>1.4 HIV prevalence and age</td>
<td>4</td>
</tr>
<tr>
<td>1.5 HIV prevalence and educational level</td>
<td>4</td>
</tr>
<tr>
<td>1.6 HIV prevalence and wealth index</td>
<td>4</td>
</tr>
<tr>
<td>1.7 HIV prevalence and area of residence</td>
<td>4</td>
</tr>
<tr>
<td>1.8 HIV prevalence and marital status</td>
<td>5</td>
</tr>
<tr>
<td>1.9 Problem statement</td>
<td>5</td>
</tr>
<tr>
<td>1.10 Research questions</td>
<td>5</td>
</tr>
<tr>
<td>1.11 Hypothesis</td>
<td>6</td>
</tr>
<tr>
<td>1.12 Objectives of the study</td>
<td>6</td>
</tr>
<tr>
<td>1.12.1 Main Objective</td>
<td>6</td>
</tr>
<tr>
<td>1.12.2 Specific Objectives</td>
<td>6</td>
</tr>
<tr>
<td>1.13 Significance and Justification of the Study</td>
<td>7</td>
</tr>
<tr>
<td>1.14 Scope of the Study</td>
<td>7</td>
</tr>
<tr>
<td>1.15 Limitations of the study</td>
<td>7</td>
</tr>
<tr>
<td><strong>2 Literature review</strong></td>
<td>8</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Literature</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Overview of the literature</td>
<td>11</td>
</tr>
<tr>
<td><strong>3 Methodology</strong></td>
<td>12</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>12</td>
</tr>
<tr>
<td>3.2 The statistical Model</td>
<td>12</td>
</tr>
<tr>
<td>3.2.1 The Log Odds Ratio Transformation</td>
<td>13</td>
</tr>
<tr>
<td>3.2.2 Odds ratios and logistic regression</td>
<td>13</td>
</tr>
<tr>
<td>3.2.3 Logistic regression</td>
<td>13</td>
</tr>
<tr>
<td>Model Assumptions</td>
<td>13</td>
</tr>
</tbody>
</table>
3.2.4 Binomial distribution

4 Results and interpretation

4.1 Introduction
4.2 Regression Output
4.3 Age
4.4 Marital status
4.5 Gender
4.6 Wealth status
4.7 Place of residence
4.8 Education attainment

5 Summary, conclusion, and policy recommendations
Figures and Tables

Figures

Figure 1. Trend of new infection ................................................................. 3

Tables

Table 1. Logistic Regression Output ............................................................. 18
Table 2. HIV prevalence and age ................................................................. 19
Table 3. HIV prevalence and Marital status ................................................. 19
Table 4. HIV prevalence Gender ................................................................. 19
Table 5. Wealth Status ................................................................................. 20
Table 6. HIV prevalence and Place of Residence ........................................... 20
Table 7. Education Attainment ................................................................. 20
Acknowledgments

I thank The University of Nairobi especially the school of Mathematics for giving me the opportunity to study a post graduate course in Biometry. I also acknowledge my esteemed supervisor Dr.Idah Orowe for guiding me through the entire project and report writing. Forever grateful
List of abbreviations

- AIDS- Acquired Immuno-Deficiency Syndrome
- ART - Antiretroviral Treatment
- HIV - Human Immuno-deficiency Virus
- KAIS - Kenya AIDS Indicator Survey
- PrEP - Pre-exposure Prophylaxis
- MOT - Modes of transmission
- MSM - Men who have sex with men
- KDHS - Kenya Demographic and Health Survey
1 Introduction

1.1 Background

Despite more than two decades of work in the field of HIV/AIDS prevention, global estimates of HIV infections indicated that, 36.7 million [30.8 million–42.9 million] were living with HIV at the end of 2016. Sub-Saharan Africa remains most severely affected, with nearly 1 in every 20 adults (4.9%) living with HIV and accounting for 69% of the people living with HIV worldwide. Kenya as one of the Sub-Saharan countries and severely affected, it declared HIV/AIDS as a national disaster in 1999, since then, Kenya has witnessed an upsurge in behavior change campaigns to prevent the disease from causing future destruction on the economy. Recent statistics has indicated that there are 62,000 new infections by the end of 2016, with the highest statics recorded in Nyanza part of Kenya.

The first case of HIV in Kenya was detected in 1984. By the mid-1990s, HIV was one of the major causes of illness in the country, putting huge demands on the healthcare system as well as the economy. In 1996, 10.5% of Kenyans were living with HIV, although prevalence has almost halved since then, standing at 5.9% by 2015. This progress is mainly due to the rapid scaling up of HIV treatment and care. In 2016, 64% of people living with HIV were on treatment, 51% of whom were virally suppressed.

1.2 HIV prevalence in Kenya

Kenya’s HIV epidemic is driven by sexual transmission which accounts for 93.7% of all new infections and is generalized, meaning it affects all sections of the population including children, young people, adults, women and men. As of 2015, 660,000 children were recorded as being orphaned by AIDS. However, a disproportionate number of new infections happen among people from key populations. According to the MOT (2008), although these populations represent less than 2% of the general population, they contribute 30% of all new HIV infections. Key populations in Kenya include sex workers, men who have sex with men (MSM) and people who inject drugs.

Geographic location is also a factor, with 65% of all new infections occurring in nine out of the country’s 47 counties – mainly on the west coast of Kenya. In particular, new HIV infections in major cities Nairobi and Mombasa increased by more than 50% (from a collective total of 4,707 in 2013 to 7,145 in 2015). As a result, HIV prevalence ranges from 0.1% in Wajir to 25.4% in Homa Bay.
1.3 HIV prevalence and gender

HIV epidemic exhibits extreme gender disparity. The prevalence among women is at 7.6% and 5.6% among men. In 2016, there were 910,000 women living with HIV out of the total 1.6 million individuals with HIV [UNAIDS (2018)]. Women in Kenya are still discriminated when it comes to access to education, health care and employment. Men usually dominate sexual relationships hence limiting the ability of women to choose safe sex practices.
1.4 HIV prevalence and age

51% of all new HIV infections in Kenya in 2015 occurred among adolescents and young people (aged 15-24 years), a huge rise from 29% in 2013. Young women are twice more likely to contract HIV as compared to their male counterparts and accounted for 33% of all new infections as compared to 16% by the males. This high HIV prevalence could be attributed to a number of factors such as incorrect perception of HIV risk and unprotected sex under the influence of alcohol and drugs. Sexual violence also increases young people’s vulnerability to HIV. This particularly affects young Kenyan women who are three times more likely to be exposed to sexual violence compared to young Kenyan men.

1.5 HIV prevalence and educational level

Education is an important factor in improving population health. More education makes individuals less vulnerable and more likely to practice safe sex. Understanding of the effects of formal education on HIV/AIDS infection in Kenya is complex because different research offers different, seemingly contradictory results and explanations of what exactly are the schooling effects on HIV/AIDS and the causal mechanisms driving those effects. Strong correlation between education level and knowledge of HIV/AIDS in Kenya was found by Selemani Mwamwenda, 2014. Earlier years into HIV pandemic, formal education was identified as a risk factor, Twenty years later; education began to shift from risk to social vaccine.

1.6 HIV prevalence and wealth index

The relationship between HIV infection and poverty has attracted considerable research attention in recent years, but the relationship is rather complex and findings from existing studies remain inconclusive. While some argue that poverty increases vulnerability, existing data from Kenya largely support the view that wealthier men and women, especially rural residents, have higher prevalence of HIV. Statistical correlations of epidemiological and socioeconomic data show that in many African countries, the prevalence of HIV infection correlates directly with wealth. For example, Shelton et al. illustrated a strong positive relationship between household wealth and HIV infection prevalence in the United Republic of Tanzania. Chin, after analyzing data from Kenya, also showed that national HIV prevalence rates appeared to correlate directly with national income across sub-Saharan Africa. Peter Piot et al. showed that in African countries HIV infection rates correlate not only with wealth, but also with income inequality.
1.7 HIV prevalence and area of residence

Urban poor in Kenya have a significantly higher risk of acquiring HIV infection than their urban non-poor counterparts. Poverty on the other hand is associated with a significantly lower risk among rural residents. Interestingly urban/rural differentials are observed in poverty risk factor by key socio-cultural and demographic characteristics, including gender disparities and ethnicity. For example, the rich-poor gap among urban residents is wider for women, with poor urban women being particularly vulnerable [Magadi (2013)].

1.8 HIV prevalence and marital status

Marriage can be a predisposing factor to HIV infection. This is because of low condom use and trust inherent within marriage. Having multiple concurrent sexual partners (having more than one partner during the same time period) has a major role in increasing the HIV infections [Halperin and Epstein (2004)]. There exists a strong association between high HIV risk and polygamy. In 2004 among married people, seven percent of those in monogamous relationships were HIV-positive, but the rate reached 11 percent among those in polygamous relationships [demographic health surveys (2004)]. HIV prevalence is highest among men and women who have ever been widowed at 19.2% and 20.3% respectively [KAIS, 2012].

1.9 Problem statement

There are a number of reasons why HIV in Africa, especially sub-Saharan Africa, has become such a serious problem. The government of Kenya and other concerned bodies have been doing a lot of activities in prevention and controlling the prevalence of HIV/AIDS since the first incidence in 1984. However, due to the wide range of differences and diversity of social groups in the country together with some cultural and socioeconomic factors and other hindrance like budgetary constraints, the effort of prevention and controlling of the epidemic hasn’t gone as expected. Therefore, the prevalence rate of the epidemic remains high in the country.

1.10 Research questions

The study seeks to answer the following questions;

- Which underlying factors are responsible for the high prevalence of HIV/AIDS in Kenya?
- Do these factors have the same effect on HIV prevalence among male and female individuals?
1.11 Hypothesis

Based on the theory and background research, I postulate the following hypothesis for further investigation;

• The probability of having HIV is higher for the less educated individuals as compared to the more educated individuals.
• An urban resident has a higher risk of HIV infection as compared to a rural counterpart.
• The women have a greater risk of HIV infection as compared to men.
• Those who are not married have a greater risk as compared to the married.
• Economically poor individuals have higher likelihood of acquiring HIV/AIDS than wealthier individuals.
• Younger age groups are more vulnerable to HIV infection than older ones.

1.12 Objectives of the study

1.12.1 Main Objective

The aim of this paper will be to identify which factors have a significant effect on the prevalence of HIV/AIDS in Kenya. This will help to identify and to figure out whether these factors play a role in maintaining the high prevalence of HIV/AIDS in the county. These will also help policy makers working on HIV prevention; understand on which groups they should focus on to mitigate the prevalence of the epidemic. There is need to tailor the prevention programs to the needs of each population group.

1.12.2 Specific Objectives

• To apply binary logistic regression to determine factors that have an influence on HIV/AIDS prevalence Kenya.
• To determine if these factors have the same effect on HIV prevalence among men and women.
• To build regression models showing the extent of the effect of this factors on HIV prevalence among men and women.
1.13 Significance and Justification of the Study

Knowledge on HIV/AIDS prevalence and its determinants provides useful information that can be used in designing the best preventive measures to curb its further spread. The best preventive measures then lead to reduction in the number of people with HIV/AIDS thereby ensuring an active and healthy population. This will be in line towards achieving the vision 2030.

1.14 Scope of the Study

The study encompasses the entire population of Kenya as a Country.

1.15 Limitations of the study

The use of models to determine the relationship between HIV prevalence and demographic and socioeconomic factors may not be very accurate. This is because most models do not capture the complex interaction between these factors.
2 Literature review

2.1 Introduction

Thousands of books, journals and material on have been written on HIV/AIDS. This paper will consider some of the papers that have been written on this subject.

2.2 Literature

Kenya reported its first AIDS case in 1984 [NJUEial (2000) AIDS and STI Control Programme (NASCOP)]. AIDS is caused by Human Immune Deficiency Virus (HIV) that impairs the immune system, making the body susceptible to and unable to fight opportunistic diseases that lead to death through secondary infections.

In 1987, Kenya reported 1,299 HIV cases. In 1990, there were 16,150 confirmed cases. In 1991, this rose to 25,702 and over 50,000 cases were confirmed in 1994. In September 1997, over 76,000 deaths were reported. There were over 1.32 million people already infected by HIV [Amuyunzu-Nyamongo (2001)]. The AIDS deaths in 1999 were over 180,000. Cumulative number of orphans by that year was 730,000 (UNAIDS, 2000). With a population of over 30 million, in year 2000, adults living with HIV/AIDS were over 2 million where 1.1 million women and 78,000 children were living with HIV/AIDS. People living with AIDS in Kenya translate to about 9.3 per cent of the total population.

Premarital sexual activity is very common and starts early in life. The breakdown in traditional family systems, urbanization and the influence of mass media are some of the factors contributing to increased sexual activity. The high level of sexual activity, which is often unprotected, is associated with HIV/AIDS, pregnancy and unsafe abortions, economic hardships and school dropouts. There are variations however based on residence, level of education and other socioeconomic status (Muganda et al., 2003).

The prevalence rate of HIV is lower in rural areas, where about 80 per cent of the total population lives than in urban areas. Still, the greatest burden of HIV infection is in the rural population. Urban residents have a significantly higher risk of HIV infection (10 per cent) than rural adult residents (6 per cent). Prevalence in urban women is 12 per cent as opposed to with less than 8 per cent of rural women (NCPD et al., 2003).

In Kenya heterosexual relations comprises of 75 per cent of all HIV transmissions. The peak ages for AIDS cases are 20-29 years for females and 30-39 for males. This is the
most economically productive group of the population and the deaths accruing from AIDS has an important economic burden((Njue, 2000; NASCOP, 2005). This is the age when investments in education are just beginning to pay off. The deaths also have important consequences for children since most people in these age groups are raising children. The worst impact is an increase in the number of orphans.

Women are more likely to be infected at a younger age than men. The high female to male HIV positive ratios in Kenya in ages 15-49 could mainly be because of the differential rates of transmission or susceptibility to infection.

Risky sexual behaviour has a correlation with age(Carael, 1995). Increase in single youth’s numbers accompanied by a decrease in the average age of sexual debut and rising levels of pre-marital sex could also explain the risky behaviour among the youth(Brown and Xenos, 1994 C.F Njue, 2000). Data show that the mean age for first sexual intercourse in Kenya is at 13 and 14 for girls and boys respectively. At age 19 years, 92.8 per cent of teenagers have initiated sex(Magadi, 1996).

HIV affects young women in Kenya disproportionately. During difficult economic conditions, young people have sexual relations with older partners who give them money or gifts in exchange for sex(UNICEF/UNAIDS, 2005).

Education is the main method through which Western ideas; norms and beliefs replace traditional ones in developing countries. In Sub Saharan Africa, the existence of indigenous and modern ideas and practices side by side create confusion and conflict (Gage, 1998)-Rapid urbanisation, westernisation and widespread availability of information through global news media, the World Wide Web and Internet coupled with education present a state of transition among the population.

Statistics South Africa (2014) released a statistical trend on a number of factors observed in the trend of population in South Africa. In the summary about HIV/AIDS, the journal states that the rate of HIV prevalence was about 10.2 percent of South African population.

The journal states that projection of the country’s demography is majorly modified by the death arising from AIDS and HIV infections that destabilize fertility in the country. These have contributed to change in life expectancy of the country.

On average, someone with HIV stays about 10.5 years then he/she dies. Female of age 15-49 years have 1.5 times more chances of being infected with HIV as compared to male counterparts.
Between 2002 and 2014, approximately 1.42 million people were HIV positive in South Africa. About a fifth of South African females were tested HIV positive while in their reproductive age bracket.

The Eastern and South African Journal (2016) updated on the websites about the HIV/AIDS trends in the region. The journal has identified the region as one hit hard by HIV/AIDS. It is stated that of 6.2% of the world’s population, which is about 19 million has been infected with the virus.

In 2015 only, about 960 thousand people were infected in the African region, of which 40% were South African and the rest were from other seven countries namely: Kenya Uganda, Tanzania, Ethiopia, Malawi, Zambia and Mozambique.

In the finding, women are the ones most infected with HIV/AIDS. The writer also tried to categorize people based on magnitude of infections. It was highest in women aged 15-24 years averagely by 3.7% unlike in men of the same age which was 1.8%. There was variation from one country to another. Some of the factors that contributed to the infections are: prostitution, homosexual acts, use of drugs by injection methods and having sex out of marital institutions.

The author also tried to explain the key causes of infections among the Eastern and Southern African population.

Some initiatives were in implementation process to control HIV infection. Counselling after testing for HIV is done in most of the countries. It showed that male and female individuals tested for and got the results for HIV was from 4% to 60% as in Madagascar and Botswana respectively. Kenya has intensified innovativeness to make people embrace testing for HIV. The government introduced self-testing kit in 2015. The kit was being circulated from door to door. This led to large increase in number of HIV testing.

Most of the countries in the region have tried to intensify prevention programs against spread of HIV. Zimbabwe, Malawi, Ethiopia and Swaziland tried, in 2015, on revitalizing prevention programs nationally. Kenya, South Africa and Zimbabwe had meetings in the same year to discuss on the way forward to eliminate HIV spread and infections.

Some of the means agreed to be used in fighting the spread of virus were: preventing transmission from mother to child. This contributed up to 66% decline in infections among children of age 0-14 years.

Most countries in the region also embraced voluntary male circumcision medically implemented. By doing this, it was estimated that about 3,400,000 people could be secured from being infected besides saving not less than 16.5 billion dollars that could have been
used in treatments only. This exercise led to at least 80% of adult men in each country being circumcised. The other methods implemented were ART and PrEP.

Lastly, the journal has tried to explain some barriers to prevention of HIV. Economic instability to facilitate treatment leads to retardation in effort to curb the spread of the virus. Most of the funding comes from international donors. There are also issues of stigmatization and discrimination among family and community members which discourage HIV victims from looking for medication. The most affected are women and girls who are discriminated on access to education.

Lastly, legal barriers discriminate certain form of sexual orientations. For instance, sex workers and homosexuals may not be free to go for condoms and HIV treatments just for fear of being jailed in some countries like Mozambique and Seychelles. These discourage the victims of such orientation from looking for medication and preventive measures and tools to spread of HIV (Eastern and Southern Africa, 2016).

### 2.3 Overview of the literature

A review of the literature reveals that formal education level, age, place of residence, gender, marital status and wealth index influences the prevalence of HIV. However, none of the studies captures the magnitude and the direction of this influence. This study seeks to fill this gap.
3 Methodology

3.1 Introduction

The purpose of this study is to identify the underlying determinants which have a significant effect on the transmission of HIV/AIDS in Kenya by using logistic regression techniques. In this chapter, the theoretical background and the method behind the analysis of the data will be presented.

3.2 The statistical Model

Due to the binary nature of the outcome variable in this study, being HIV positive or negative, a binary logistic regression model will be used. One of the common applications of logistic regression is to determine the chance of the occurrence of a particular outcome of the response variable on the basis of explanatory variables by fitting a given data to logit function. Based on the number of categories of the outcome variable, a logistic regression model can be classified as binary or multinomial. The outcome variables which are binary in nature are classified under binary logistic regression whereas outcome variables which have more than one category are categorized under multinomial. In multiple regressions, a mathematical model of a set of explanatory variables is used to predict the mean of a continuous dependent variable. In logistic regression, a mathematical model of a set of explanatory variables is used to predict a logit transformation of the dependent variable. Suppose the numerical values of 0 and 1 are assigned to the two outcomes of a binary variable. Often, the 0 represents a negative response and the 1 represents a positive response. The mean of this variable will be the proportion of positive responses. If \( p \) is the proportion of observations with an outcome of 1, then \((1-p)\) is the probability of a outcome of 0. The ratio \( \frac{p}{(1-p)} \) is the odds and its logarithm is called logit or log odds. The logit transformation is written as follows;

\[
I = \logit(p) = \ln \left( \frac{p}{1-p} \right)
\]
3.2.1 The Log Odds Ratio Transformation

To compare two proportions, such as that of males versus females, the difference between two log odds can be calculated.

\[ l_1 - l_2 = \logit(p_1) - \logit(p_2) \]

\[ = \ln\left(\frac{p_1}{1-p_1}\right) - \ln\left(\frac{p_2}{1-p_2}\right) \]

\[ = \ln(OR_{12}) \]  

3.2.2 Odds ratios and logistic regression

In logistic regression, the regression coefficient \((\beta_1)\) is the estimated increase in the log odds of the response variable per unit increase in the value of the independent variable. Meaning, the exponential function of the regression coefficient \(e^{\beta_1}\) is the odds ratio associated with a one-unit increase in the independent variable.

3.2.3 Logistic regression

Binary Logistic regression for continuous explanatory variables

A binary response \(Y_j\) and an explanatory variable \(X_{ij}, \ i=1,2,\ldots,m \) and \(j=1,2,\ldots,n\), where: m is the number of variables in the model n is the number of observations. Let \(\pi_j = P(X_{ij})\) represent the chance of occurrence of an event when \(X_{ij}\) takes the values \(x_{ij}\).

Model Assumptions

1. The independent variable need not be normally distributed.

2. The model do not assume a linear relationship between the response and the independent variable. The linear relationship exists between the logit of the response and explanatory variables.

3. The error term is independent.

4. Independent variables lack co-linearity. The resulting models are therefore easier to interpret.

The problem with a linear model is that the probability model \(E(Y) = X\beta\) is used to approximate a probability value, \(\pi_j = P(Y_j = 1)\) within the interval 0 and 1, while \(E(Y_j)\) is
not so constrained. Hence, we use the logit transformation where the transformed quantity $\ln \left( \frac{\pi_j}{1 - \pi_j} \right)$ exists in the interval $(\infty, \infty)$. The model is as follows: \[ \text{Logit}(\pi_j) = \ln \left( \frac{\pi_j}{1 - \pi_j} \right) = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_m X_{mj} \tag{5} \]

Through algebraic manipulation \[ \pi_j = \frac{\exp(\beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_m X_{mj})}{1 + \exp(\beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_m X_{mj})} \tag{6} \]

Where the parameter $\beta_i$ represents the coefficient of parameter to be estimated.

In the logistic regression model, maximum likelihood estimation technique is used for parameter estimation. The aim of logistic regression is to estimate the $K+1$ unknown parameters $\beta$. Maximum likelihood estimation is about finding the set of parameters for which the chance of the observed data is the greatest.

Maximum likelihood equation is derived from the probability distribution of the dependent variable. Each $y_i$ represents a binomial count in the $i^{th}$ population, the joint probability function of $Y$ is \[ f(y|\beta) = \pi \frac{n_i!}{y_i!(n_i-y_i)!} \pi^{y_i}(1-\pi)^{n_i-y_i} \tag{7} \]

For each population, there are $\binom{n_i}{y_i}$ ways to arrange $y_i$ successful attempts from among $n_i$ trials. When the probability of $y_i$ successes is $\pi$, the probability of $n_i - y_i$ failures becomes $(1-\pi)^{n_i-y_i}$.

The joint probability density function gives the values of $y$ as a function of known fixed values of $\beta$. The likelihood function has the same form as the probability density function. The likelihood function gives the values of $\beta$ in terms of fixed, values of $y$, hence \[ L(\frac{y}{\beta}) = \pi \frac{n_i!}{y_i!(n_i-y_i)!} \pi^{y_i}(1-\pi)^{n_i-y_i} \tag{8} \]

$\beta_0$ = the intercept term
$\beta_1, \beta_2, \ldots, \beta_m$ = Regression coefficients.
$X$’s are the explanatory variables.

Notice that in equation (1), although the regression model is linear on the right side, the left side is a non-linear function of the response variable. This function is called the logit link function. As mentioned above the usual least squares methods cannot be used to
estimate the parameters. Instead the maximum likelihood method is used to obtain these estimates and since the regression model is non-linear, an iterative algorithm is needed for parameter estimation.

**Binary Logistic regression model for categorical predictors**

Logistic regression can have categorical predictors or continuous explanatory variables. Since I am going to use categorical predictors in this study, logistic regression for categorical predictors will be presented below. Suppose the model has a binary response $Y$ and $m$ predictors, $i = 1, 2, \ldots, m$. The predictors may have more than two levels. $X_{ij}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n, r = 1, 2, \ldots, k_i - 1$

Where $m$ is the number of variables included in the model. $n$ is the number of observations. $X_{ij}$ refers to the $r^{th}$ level of a factor. One level of each factor will be taken as a reference category; therefore the model will have $k_i - 1$ dummies.

$$P(Y = 1) = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_{k_i-1} X_{k_i-1j} + \beta_1 X_{1j}^2 + \beta_2 X_{2j}^2 + \ldots + \beta_{k_i-1} X_{k_i-1j}^2 + \ldots + \beta_1 X_{mj} + \beta_2 X_{mj}^2 + \ldots + \beta_{k_m-1} X_{kmj}$$

(9)

Using the above equation, a general logistic regression models will be built based on the KAIS 2012 data and will be analyzed using the SPSS software. The general model include all the selected variables and will capture the effects of each explanatory variable on the odds of being infected by the virus. The general model can be expressed as:

$$\text{Logit}(\pi_j) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Educ} + \beta_3 \text{Res} + \beta_4 \text{Mart} + \beta_5 \text{Weal}$$

(10)

### 3.2.4 Binomial distribution

$$f(y, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y},$$

(11)

where $n$ is the size of the sample, $y$ is the random variable, $\theta$ is the probability of successes, $(1 - \theta)$ is the probability of failure.

$f(y, \theta)$ can be written as:

$$f(y, \theta) = \exp[\ln\binom{n}{y} + y\ln\theta + (n - y)\ln(1 - \theta)]$$

(12)

Thus

$$= \exp[\ln\binom{n}{y} + y\ln\theta + n\ln(1 - \theta) - y\ln(1 - \theta)]$$

(13)

$$= \exp[\ln\binom{n}{y} + y\ln(\frac{\theta}{1 - \theta}) + n\ln(1 - \theta)]$$

(14)

On rearranging, we get

$$= \exp[y\ln(\frac{\theta}{1 - \theta}) + n\ln(1 - \theta) + \ln\binom{n}{y}]$$

(15)

where

$$a(y) = y; b(\theta) = \ln(\frac{\theta}{1 - \theta}); c(\theta) = n\ln(1 - \theta); d(y) = \ln\binom{n}{y}$$

(16)
**Interpretation of the coefficients of Logistic regression model for categorical predictors**

In logistic regression, one level of the dependent and independent variable should be selected as a reference level. The impact of explanatory variables on the response of the dependent variable is given by the coefficients of regression. The sign of the regression coefficient tells us whether the independent variables increased or decreased the likelihood of the outcome as compared to the reference level. A positive regression coefficient indicates that the independent variables increase the probability of the outcome and the opposite effect will be showed when the regression coefficient is negative. The chance of the outcome of the response variable is determined by the magnitude of the regression coefficient of the given independent variables. Big regression coefficient means the independent variables affect the chance of the outcomes occurring strongly; on the other hand small coefficient has opposite effect, providing that the reference category sets to zero. The effect of the explanatory variable on the outcome of the dependent variables explained by 's unit that means the intercept term \(( \beta_0 )\) can be interpreted as the value of the log-odds of success outcome when all the risk factors \((X's)\) are set to their reference level. To interpret the regression calculating odds ratio will be very useful because it is reasonable and comparatively easy to interpret and understand the result of the logistic regression, as compared with the log-odds. The odds ratio (OR) gives us the strength of association between an independent variable and an outcome dependent variable. The odds ratio can be easily calculated using different statistical software, and its value lies in the interval between 0 and \( \infty \). In this paper, SPSS software will be used for calculating the odds ratio and estimation of regression coefficients. The odds ratio of a logistic regression for categorical predictors can be interpreted as the change of the level of a given categorical explanatory variables as compared to that of reference level, provided that all other levels assigned to zero. The software gives a p-value of the Wald test, and the estimated regression coefficients. Based on the values of the odds ratio it is possible to interpret the result. If the odds ratio is greater than one the likelihood of the occurrence of an event is high, whereas if odds ratio is less than one then opposite effect is true. If the odds ratio is equal to one then the occurrence of the two outcomes are equally likely.

**Definition of model variables**

Variables to be included in this study are selected from 2012 Kenya Aids Index Survey.

**Dependent variable**

The endogenous variable in this study is the individuals’ serostatus of HIV; we give the value 0 if the respondent HIV testing result is negative and 1 if the participant’s HIV testing result is positive.
Independent variables

The independent variables that are used in the study were gender, age, educational level, wealth index, residence and marital status. Each variable has its own levels. Age is one of the Demographic variables influencing the HIV epidemic. In this study the variable age is Categorized into 3 levels. Before editing it had 10 levels in the KAIS 2012 data set for female and male respondents. Here I am interested to know which age groups are more vulnerable to HIV epidemic, young aged (15-24), middle aged (25-34) and older aged (35-44). For the marital status of the individuals, we have made some modification on the variables. According to the KAIS 2102 the variable has 5 levels but now it is reduced to 3 for the sake of convenience. KAIS 2012 classified this variable in detail, like the one who had wife/husband previously was classified as divorced and widowed, but here they are merged together. Also those married have been classified as in monogamous and polygamous relationship but for the sake of this study the two will be merged. Hence, the variables categorized in this way, it will have three main levels. The other determinant factor that affects the prevalence of HIV epidemic for both women and men individuals is the wealth index, it has three levels.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>VARIABLE LABEL</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age group</td>
<td>1=15-24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=25-34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3=35-44</td>
</tr>
<tr>
<td>Marts</td>
<td>Marital status</td>
<td>1=Never in a union</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=Married</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3=Widowed/separated/divorced</td>
</tr>
<tr>
<td>Gender</td>
<td>Sex of the respondents</td>
<td>1=Male</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=Female</td>
</tr>
<tr>
<td>Wealth</td>
<td>Wealth index</td>
<td>1=Poorer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=Middle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3=Rich</td>
</tr>
<tr>
<td>Place</td>
<td>Place of residence</td>
<td>1=Rural</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=Urban</td>
</tr>
<tr>
<td>Education</td>
<td>Educational attainment</td>
<td>1=No education</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=Primary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3=Secondary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4=Higher</td>
</tr>
</tbody>
</table>
4 Results and interpretation

4.1 Introduction

This chapter gives the results of the findings obtained by estimating the model and the interpretations of the results. The output of the logistic regression used to measure the relationship between the prevalence of HIV and the key independent variables will be discussed at 5% level of significance.

4.2 Regression Output

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>β</td>
<td>sig</td>
</tr>
<tr>
<td>Age</td>
<td>0.0052</td>
<td>0.000</td>
</tr>
<tr>
<td>Marts</td>
<td>0.0360</td>
<td>0.010</td>
</tr>
<tr>
<td>Wealth</td>
<td>-0.991</td>
<td>0.487</td>
</tr>
<tr>
<td>Edu</td>
<td>-0.803</td>
<td>0.622</td>
</tr>
<tr>
<td>Res</td>
<td>-0.0125</td>
<td>0.016</td>
</tr>
<tr>
<td>Cons</td>
<td>-0.559</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Male Regression Model

\[ \text{Logit}(\pi_j) = -0.559 + 0.0052\text{Age} - 0.0125\text{Res} + 0.0360\text{Marts} \] \hspace{1cm} (17)

Female Regression Model

\[ \text{Logit}(\pi_j) = -0.056 + 0.3357\text{Age} - 0.006\text{Educ} - 0.2756\text{Res} + 0.1250\text{Marts} - 0.1129\text{Wealth} \] \hspace{1cm} (18)

4.3 Age

The age group 15-24 was used as the reference group. The odds of being infected by HIV/AIDS in the middle age groups are significantly less as compared to young age group. As shown in the model, the odds of being infected with HIV/AIDS in the age group 25-34 was about 0.06 and 0.34 less for male and female respectively as compared to the very young age groups (15-24). However, unlike for male respondents, the odds of acquiring of HIV for female respondents in the age group 35-44 were significantly high at 1.5 times that of their younger counterparts.
### Table 2. HIV prevalence and age

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Male β</th>
<th>sig</th>
<th>Exp(β)</th>
<th>Female β</th>
<th>sig</th>
<th>Exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>15-24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2.784</td>
<td>0.000</td>
<td>0.062</td>
<td>-1.081</td>
<td>0.000</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>25-34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.756</td>
<td>0.001</td>
<td>0.047</td>
<td>0.414</td>
<td>0.000</td>
<td>1.513</td>
</tr>
<tr>
<td></td>
<td>35-44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4.4 Marital status

The coefficient estimates of the marital status model show that the effect of marital status on the odds of being infected by the virus is statistically significant for male both male and male respondents. The likelihood of being infected by the virus was increased by marriage and is even higher among the widowed and divorced. For example widowed or divorced women carry almost two times the risk of infection as compared to their unmarried counterparts.

### Table 3. HIV prevalence and Marital status

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Male β</th>
<th>sig</th>
<th>Exp(β)</th>
<th>Female β</th>
<th>sig</th>
<th>Exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marts</td>
<td>N/Marr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Married</td>
<td>-0.173</td>
<td>0.000</td>
<td>1.188</td>
<td>0.358</td>
<td>0.000</td>
<td>1.430</td>
</tr>
<tr>
<td></td>
<td>Wid/Div</td>
<td>0.595</td>
<td>0.001</td>
<td>1.813</td>
<td>0.667</td>
<td>0.000</td>
<td>1.948</td>
</tr>
</tbody>
</table>

### 4.5 Gender

Sex of respondents had also a significant effect on the prevalence of HIV. In this regard the odds of acquiring HIV by the women are increased by 2.8 times as compared to the males.

### Table 4. HIV prevalence Gender

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Male β</th>
<th>sig</th>
<th>Exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1.052</td>
<td>0.009</td>
<td>2.874</td>
</tr>
</tbody>
</table>

### 4.6 Wealth status

The regression output a showed that the likelihood of women being infected by the virus was overall statistically significantly affected by the wealth of the respondent. The odds of being HIV
positive for middle and richer women were about 1.102 and 1.107 times higher than poorer women, respectively. The odds of being infected were insignificant among men.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Male β</th>
<th>sig</th>
<th>Exp(β)</th>
<th>Female β</th>
<th>sig</th>
<th>Exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wealth</td>
<td>Poorer</td>
<td>-0.668</td>
<td>0.059</td>
<td>0.512</td>
<td>0.098</td>
<td>0.000</td>
<td>1.102</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>0.352</td>
<td>0.582</td>
<td>1.274</td>
<td>0.102</td>
<td>0.014</td>
<td>1.107</td>
</tr>
</tbody>
</table>

### 4.7 Place of residence

The risk of being infected with the HIV virus was significantly higher in the urban areas as compared to the residents of the rural areas for both men and women.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Male β</th>
<th>sig</th>
<th>Exp(β)</th>
<th>Female β</th>
<th>sig</th>
<th>Exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
<td>Urban</td>
<td>-0.012</td>
<td>0.000</td>
<td>0.988</td>
<td>-0.118</td>
<td>0.000</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>-0.012</td>
<td>0.000</td>
<td>0.988</td>
<td>-0.118</td>
<td>0.000</td>
<td>0.888</td>
</tr>
</tbody>
</table>

### 4.8 Education attainment

The model regarding educational level revealed that there was significant variation in the likelihood of getting HIV AIDS among women who attained some educational level and women who had no education at all. However, this variation turned to insignificant for men. For women, the risk of infection is higher among those with some education than those with no education at all. For example the risk is 1.3 times higher among those educated to the primary level as compared to those who didn’t go to school.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Male β</th>
<th>sig</th>
<th>Exp(β)</th>
<th>Female β</th>
<th>sig</th>
<th>Exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>No Edu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary</td>
<td>-0.347</td>
<td>0.559</td>
<td>0.707</td>
<td>0.297</td>
<td>0.001</td>
<td>1.345</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>-0.937</td>
<td>0.113</td>
<td>0.393</td>
<td>0.274</td>
<td>0.000</td>
<td>1.315</td>
</tr>
<tr>
<td></td>
<td>Higher</td>
<td>0.134</td>
<td>0.791</td>
<td>1.114</td>
<td>0.170</td>
<td>0.002</td>
<td>1.185</td>
</tr>
</tbody>
</table>
5 The Summary, The conclusion, and policy recommendations

5.1 Introduction

In this section the summary, the conclusion and recommendation of the study are presented.

5.2 Summary

This study examined the factors that have a significant effect on the prevalence of HIV/AIDS. Hence in the next section all the explanatory variables which have been identified to have a significant effect on the prevalence of HIV/AIDS at the national level will be discussed.

As we mentioned in the theoretical background section the youngest age group is expected to rush to sex and practice unsafe sex. Parallel to these as stated on the hypothesis section these age group members are more vulnerable to HIV infection than older age groups. They have low awareness about the epidemic and ways to protect themselves, and they may not have access to protective measures such as condoms, or are not able to afford them; hence they are easily exposed to the virus. 51% of all new HIV infections in Kenya in 2015 occurred among adolescents and young people (aged 15-24 years), a huge rise from 29% in 2013 (Kenya AIDS Response Progress Report 2016). This is consistent with the logistic regression result which showed that, as we expected, the young age group (15-24) was found to be at the highest risk of contracting the virus as compared to middle and older age groups for both sexes.

Marital status has been found to have a significant effect on the prevalence of HIV. Some scholars have suggested that marital status has a positive association with the prevalence of HIV because of low condom use and trust inherent in marriage [Das et al., 2004]. From the logistic regression, the result showed that the odds of being HIV positive are significantly higher in the married and widowed/separated/divorced individuals than in those who have never married among both sexes and age groups, hence the result is inconsistent with the stated the hypothesis.

Education has two opposite effects on the prevalence of HIV. For example education and vulnerability to HIV among the rich people were highly correlated due to greater personal autonomy and spatial migration. On the other hand, the prevalence rate is expected to be lower in this group through better awareness about the virus and ease of access to condoms and other preventive measures. Uneducated individuals could have a higher exposure to the virus due to their low income and inadequate information about the way to prevent the epidemic. The regression result shows that the level of education had insignificant effect on the prevalence of HIV/AIDS among male respondents. However, among women, the odds of being HIV positive are significantly higher in primary, secondary and higher educational level than who have no schooling, which is exact the opposite of what I expected.
Higher prevalence rate of HIV/AIDS is expected in urban areas compared to rural areas. This is because in urban areas there is a high population per square kilometer and hence a high prevalence of sexual networking and its related factors [De Vylder(1993)]. In our regression model, the odds of being HIV positive are significantly higher among the urban dwellers.

As discussed in the theory section, the relationship between HIV and poverty is a complex one. Poverty can increase the vulnerability of contracting HIV through migration, limited media exposure, sexual exploitation and gender inequality. On the other hand, studies have shown a strong correlation between high income and high HIV infection in African countries. Peter Piot et al showed that in African countries, HIV infection rates correlate not only with wealth but also with income inequality. The regression result showed that wealth status has an insignificant effect on the prevalence of HIV/AIDS among male individuals. This is similar to the education level model where education has no significant effect on HIV prevalence among men. Education and wealth has been shown to be highly correlated. However, contrary to our expectation, middle class and richer female respondents were found to be more likely to be HIV positive as compared to those with no education.

Women have been disproportionately affected by HIV in Kenya. Today, women constitute more than half of all people living with HIV (UNAIDS, 2017). AIDS-related illnesses remain the leading cause of death for women of reproductive age (15-44). Many factors increase women’s vulnerability to HIV acquisition, including, biological, behavioral, socioeconomic, cultural and structural risks. From our regression model, the odds of getting HIV as a woman is almost three times that of the male counterparts. This is consistent with our hypothesis. Considerable work needs to be done in biomedical, behavioral and structural strategies for HIV prevention with the aim of developing appropriate HIV prevention packages which take into account the socioeconomic and cultural context of women in Kenyan societies.

### 5.3 Conclusion

In this section the research questions which were posed in this paper will be answered by reviewing the result and discussion sections.

Age and marital status had a significant effect on HIV prevalence among female and male respondents. The findings showed that the odds of being HIV positive are higher in the age groups 15-24 and among married and divorced/widowed/separated individuals.

Unexpectedly poverty was found to have an insignificant effect on the prevalence of HIV/AIDS among male respondents. Unlike for male respondents, wealth status was revealed to have a significant effect on the prevalence of HIV/AIDS for female individuals.

Similarly, significant effects were found on the respondents, level of education, place of residence and gender. Odds of being HIV positive are significantly higher among urban dwellers and female gender.
5.4 Policy Recommendation

To bring about the desired rapid decline in AIDS deaths, there is a need for tailored preventive measures for different population groups. For example, campaigns appealing to young girls aged between 15 and 24 can be initiated in schools.
References


