MODELING OF STI PREVALENCE AMONG HIV-INFECTED ADULTS IN HIV CARE PROGRAMS IN KENYA USING LOGISTIC REGRESSION

Linda Akoth Chaba (I56/77435/2009)

Project submitted to the School of Mathematic

at The University of Nairobi,

In partial fullfillemt of the requirements for the award of degree of Master of Science in Biometry

Declaration

I Linda Akoth Chaba declare that this project report is my own work. It is being submitted for the degree of Masters of Science in Biometry In the University of Nairobi. It has not been submitted before for any degree or examination at this or any other University.

L Akoth — 2/08/2011

Linda Akoth Chaba — Date

I endorse the declaration by the candidate.

2/08/2011

Dr. Nelson Owuor — Date

Supervisor

Dedication

Dedicated to

My parents and siblings

(Who do not need to read it);

To my classmates

(Who might read it);

To my supervisor Dr. Nelson Owuor

(Who actually did);

And all statistics students and the School of Mathematics, Chiromo

(Who better).

Acknowledgement

I would like to pass my deepest gratitude to the Almighty God who has truly sustained me through the demanding and exigent moments of developing this project system by granting me peace of mind and health.

To my supervisors, who were readily on hand and willing to assist when need arose.

To my friends, whose integrating strength and encouragement propelled me through their taxing moments. It is in this regard that I would like to appreciate the support of my colleagues, whose assistance can seldom be sidelined.

Contents

List of Figures

List of Tables

Acronym

AIDS	Acquired Immune Deficiency Syndrome
AIC	Akaike Information Criterion
ART	Anti Retroviral Treatment
BV	Bacterial Vaginosis
CI	Confidence Interval
CT	Chlamydia Trachomatis
GC	Neisserian Gonorrhea
GUD	Genital Ulcer Disease
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
HSV-2	Herpes Simplex Virus Type 2
KAIS	Kenya Aids Indicator Survey
PID	Pelvic Inflammatory Disease
RPR	Rapid Plasma Reagin
RTI	Reproductive Track Infection
STI	Sexually Transmitted Infection
TPHA	Treponema Pallidum Haemagglutination Test
TV	Trichomoniasis Vaginalis
UD	Urethral Discharge
VD	Virginal Discharge
WHO	World Health Organisation

Abstract

Objectives and methods: This paper investigates risk factors associated with prevalence of sexually transmitted infections. Logistics regression is employed to determine the risk factors. Furthermore, the study validates the syndromic diagnosis of STI using etiological diagnosis as the gold standard. Sensitivity, specificity and positive predictive values is used in this in this validation process.

Results: 9.81% of HIV-infected participant had an etiologic diagnosis with majority being trichomoniasis (8.04% overall prevalence). There was much lower prevalence of gonorrhea (1.33%), chlamydia (0.36%), and syphilis (0.48%). 69% of the participants had positive serology tests for HSV-2. Among women who participated in the study, 18.6% were diagnosed with bacterial vaginosis.

Sensitivity, specificity and PPV of genital ulcer were 0.00%. Sensitivity of urethral discharge for predicting Gonorrhea or chlamydia low (20%), specificity was high at (95%) while PPV was 4%. Sensitivity, specificity and PPV of virginal discharge for predicting Gonorrhea, chlamydia or trichomoniasis in females are 36%, 76%, and 18% respectively.

The odds of having STI for male was 0.33 (95% CI : 0.20- 0.53) compared to women. The odds of having STI for the participants aged 45 years and above was found to be 0.33 (95% CI : 0.15 - 0 .71) as compared to those of age between 18 to 24 years old. In addition, those with vocational training/secondary education and post secondary education and above had odds ratio of 0.56 (95% CI : 0.39-0.82) and 0.05 (95% CI : 0.01-0.38). Those with more than one sexual partner was significantly associated with having STI, odds ration of 3.45 (95% CI: 1.43-8.30).

Conclusion: Prevalence of any STI was relatively low compared to results from other studies. Trichomoniasis was overwhelmingly the most common genital non-viral infection,while HSV-2 had the highest overall prevalence. Being a female,

younger age, less education, and more recent sexual partners were all associated with an increased risk of STI. Syndromic diagnosis of UD, VD and GU were insensitive, while their specificity were relatively good.

Chapter 1

Introduction

Historically, STIs have been overlooked in the global fight against infectious diseases; as a result they continue to drain the lives of young and old throughout the developing world. Although sexually transmitted Infections (STIs) have been causing significant morbidity and mortality for years, it is only with the advent of the human immunodeficiency virus (HIV) that STI control is now receiving higher priority in both developed and developing countries. This is because STIs increase the transmission of HIV and have similar behavioral risk factors.Globally, it is estimated that as many as 333 million new cases of curable STDs occur each year. The indisputable facts that STDs produce serious economic, social and health consequences, made more clear by their association with HIV, and that all STDs are preventable and many are curable, make it incumbent on governments, communities and donors to meet the challenge of STD prevention and control[14][25]. Numerous studies have demonstrated clearly that, as part of a comprehensive HIV prevention program, it is essential to take steps to reduce the spread of STDs. One of the ways is through STI management. Prevention of STI in developing countries and the resulting potential for reduced risk of HIV-1 acquisition and other STI-associated morbidity will in part depend on identification of risk factors for STI that are susceptible to interventions practical in resource poor communities.

This project therefore seek to employ logistics regression model in identifying the risk factors associated with STI prevalence.

1.1 Problem Statement

For HIV-infected persons, genital infections can lead to significant health problems, including serious short and long-term complications. In addition, co-infection with genital pathogens has been associated with an increase in blood plasma and HIV RNA levels in genital secretions. Thus, from the public health perspective, pre-existing genital infections may increase HIV transmission as a result of increased genital HIV shedding.

1.2 Problem Justification

STI constitutes major public health concern in both developing and developed countries. The emergence of AIDs has demanded measures aimed at control of STIs. A proper understanding of STI risk factors is necessary for proper implementation of STI control strategies. It therefore important to conduct a study to analyze risk factors associate with the spread of STI and also to assess the effectiveness of current approaches used in STI management.

1.3 Objectives

1.3.1 Primary objective

To determine factors associated with STIs Prevalence in HIV infected patient using logistic regression.

1.3.2 Secondary objectives

1. Determine the prevalence of STIs in HIV-infected patients in care.
2. Compare syndromic diagnosis vs etiological diagnosis of STIs

1.4 Hypothesis

1. Number of sexual partners,age, gender, education level,Condom use, and use of alcohol are the main factors associated with the spread of STI
2. Tricomoniasis is the most common STI among HIV infected adults.
3. Syndromic diagnosis is not an adequate detection tool for specific aetiology of STIs

Chapter 2

Literature Review

A number of methods have been used to determine risk factors associated with prevalence of STI and one of the common methods applied in these studies is Logistic Regression. The logistic regression model is one of the popular mathematical models for the analysis of binary data with applications in physical, biomedical, and behavioral sciences, among others. The feature of this model is to quantify the effects of several explanatory variables on one dichotomous outcome variable. As with all techniques, logistic regression has some important shortcomings. A common criticism is that logistic regression coefficients are based on the values of independent variables [10]. Although this may not be a problem with large datasets, the results might be biased coefficient for smaller sample sizes. Current advances such as exact logistics regressing tries to overcome the sample size restriction [15]. Another problem is multicolliniarity which can be checked by Haitovsky test (a test for multicollinearity that examines the null hypothesis that the matrix of correlations among predictor variables is singular with a determinant of zero). Another potential problem in logistic regression is the outliers that might alter the results. Sabrina and Mavidana (2002) are some of the papers where logistics regression has been used in determining risk factors of STI. David W. ,Scott T, and Stanley L. (1991) mentioned that logistic regression is one such

technique that can yield elegant results with a biologically relevant interpretation. However, the simple act of running a logistic regression package in no way guarantees that (1) the logistic model was the appropriate model to use and/or (2) the model accurately reflects the experience in the data. They therefore suggests that editors should require that authors provide some assurance that all logistic regression models have been checked for fit and adequacy before accepting inferences based on these models

2.1 Correlates of STI

One of the risk factors associated with STI prevalence is condom use. Warner Lee and colleagues reviewed studies published 1966-2004 to assess risk reduction for gonorrhea and/or chlamydia associated with male condom use. Of 45 studies identified, most found reduced risk of infection associated with condom use. Eight of 10 studies with 2 or more of these attributes reported statistically significant protective effects for condom use versus 15 of 35 studies with zero or one attribute (80% vs. 43%, P = 0.04) [23].

Of 42 eligible studies to conduct a systematic review of published literature on the association between problematic alcohol consumption and sexually transmitted diseases (STIs), 11 included specific measures of problem drinking, of which 8 found a significant association between alcohol consumption and at least 1 STI. The relationship did not appear to vary according to gender or pattern of alcohol consumption assessed (Cook).

The effect of male circumcision on the risk of acquiring STDs is difficult to assess due to the transient nature of many STDs. Further, reviews of studies of this association are prone to difficulties due to the absence of consistent case definitions [20]. Nevertheless, a review of observational studies found an association

between lack of male circumcision and increased risk of genital ulcer disease, particularly chancroid and syphilis [16]. There was no clear association between male circumcision and other STIs due to too few studies or inconsistent findings between studies. Epidemiological studies published since the recent meta-analysis have also found circumcision to be associated with a reduced risk of HIV infection. Lavreys and colleagues conducted a cohort study in 746 Kenyan trucking company employees. A significantly lower incidence of HIV infection was observed in circumcised men (2.5 per 1000 person-years) compared with uncircumcised men (5.9 per 1000 person-years) and this effect was stronger after adjusting for confounders including religion (adjusted rate ratio 0.25; 95% confidence interval 0.1-0.5)

Gender and age are also associated with increased risk for STIs. Women are at higher risk than men for most STIs, and young women are more susceptible to certain infections than older women. Due to cervical ectopy that is extremely common in adolescent females, the immature cervix of adolescent females is covered with cells that are especially susceptible to STIs such as chlamydia [5].

Multiple Partners is another factor associated with STI prevalence as highlighted in a document "Top 10 Risk Factors for Acquiring an STD [21]". Its pretty straightforward math; the more partners one has, the more likely it is that they will be exposed to an STI. Furthermore, people with multiple partners tend to choose partners with multiple partners, so each individual they are having sex with is probably more likely to have an infection than someone with whom one would choose to be monogamous. Prevalences of GC, CT, and TV were 2.6%, 3.2%, and 20.4% respectively (23.9% overall), and were similar at intervention and control sites. Baseline STI was associated with unmarried status, non-use of family planning, alcohol use, and more than one recent sexual partner, but the highest odds ratio was 1.5 [19].

2.2 Prevalence of STI

Relatively few studies have estimated the prevalence of STI among HIV-infected individuals. According to systematic review by Kalichman and colleagues, trichomoniasis was the most common (18.8% prevalence), followed by syphilis (9.5%), gonorrhea (9.5%), and chlamydia (5%); the average point prevalence of any of these STI was 16.3%. Only 2 of these studies occurred in sub-Saharan Africa Few other data are available regarding the prevalence of STI/RTI in HIV-infected persons in Kenya. From 1993-2004, results from two large cohorts of HIV-1 infected pregnant women at 32 weeks gestation show a decreasing prevalence of gonorrhea, chlamydia, and syphilis between time periods, although there was still a worrisome level of infection for this vulnerable group (Scott McClellan, personal communication, Table 2.1).

Table 2.1: Prevalence of genital infections at 32 weeks gestation in two cohorts of HIV-1 infected pregnant Kenyan women, 1993-2004

	1993-1998 cohort (n=425 women)	**1999-2004 cohort (n=468 women)**
Cervical mucopus by clinical exam	29.8%	9.6%
Vaginal discharge by clinical exam	51.2%	50.4%
Gonorrhea	6%, culture	1.7%, PCR
Chlamydia	11%, antigen detection	3.8%, PCR
Syphilis (RPR)	7%	1.8%
Trichomoniasis (wet prep)	24%	16.8%
Candida (KOH prep)	31%	30.3%
Bacterial vaginosis (Nugent score)	50%	33.1%
Genital ulcers (clinical exam)	13%	4.9% *

More recently, preliminary results from the 2007 Kenya AIDS Indicator Survey (KAIS), a nationally representative sample 15,872 individuals between the ages of 15 to 64 years, found that while syphilis prevalence was low in Kenya, with only 1.8% of KAIS participants infected, syphilis was 2.5 times more common (4.5%) among HIV-infected individuals than among those without HIV 4. Prevalence was similar between men and women and increased with age, numbers of lifetime sexual partners, years of sexual activity, and was higher in uncircumcised than circumcised men. The results of STI surveillance among HIV-infected adults entering into HIV care programs in Mozambique may have relevance to the Kenyan population. In Mozambique, a total of 498 patients (240 men and 258 women) were enrolled in a STI surveillance study in 2008 (Ron Ballard, personal communication). In this assessment,STI symptoms were reported by 20.5% of men and 63.5% of women. More urethral discharge (UD), genital ulcers or blisters, and genital warts were identified by providers on male genital exam than were reported by the male patients, and more vaginal discharge (VD) and genital warts were identified by providers on female genital exam than were reported by the female patients. Serological evidence of HSV-2 infection was present in 91.0% of all patients. Serologic testing for syphilis found that 15.2% of these HIV-infected patients were RPR and TPHA positive, much higher than the prevalence seen in the Kenyan pregnant women cohorts or KAIS, as noted above. The prevalence of gonorrhea and chlamydial infections was low (1.7% and 1.5%, respectively), with the prevalence of gonorrhea similar to what was seen in the pregnant women cohorts, and that of chlamydia higher in those cohorts. The prevalence of T. vaginalis and M. genitalium infections was high in both men (10.2% and 11.4%, respectively) and women (48.5% and 11.8%, respectively). Similar rates of STI diagnoses were detected among asymptomatic and symptomatic patients. The findings of these studies reinforce the need for better data to characterize the burden of STI in HIV-infected persons to inform prevention, care, and treatment

program planning.

2.3 Syndromic Diagnosis vs Etiological Diagnosis of STTs

2.3.1 Etiologic Diagnosis

This involves using laboratory to identify the causative agent. This approach avoids over treatment, conforms to traditional clinical training satisfying patients who feel not properly attended to and can be extended as screening for the asymptomatics. However, it requires skilled personnel and consistent supplies, treatment does not begin until results are available, it is time consuming and expensive, testing facilities are not available at primary level, some bacteria fastidious and difficult to culture (H.ducrey, C.trachomatis), lab results are often not reliable, mixed infections often overlooked and miss-treated/untreated infections can lead to complications and continued transmission.

2.3.2 Syndromic Diagnosis

This approach uses clinical algorithms based on the constellation of patient symptoms and clinical signs to determine antimicrobial therapy. Antimicrobial regimens are chosen to cover the major pathogens responsible for the STD syndrome in the specific geographic area. This approach provides a tool to manage symptomatic individuals but does not address the problems of sub-clinical or asymptomatic STIs or poor treatment-seeking behavior (for example, delay in seeking care after the onset of symptoms and self treatment) by individuals with symptoms. The main disadvantage of syndrome management is over-diagnosis and over-treatment, when several antimicrobials are administered to individuals with only one or no infection. Costs of over-diagnosis and over-treatment include that

of the antimicrobial itself as well as the difficult to quantify unintended adverse outcomes, such as the risk of adverse drug reactions, disruption of the normal flora of the host and its possible protective effect, domestic violence, and pressure for selecting resistant pathogens in the community. Syndromic approach is problem oriented (responds to patient's symptoms), highly sensitive and does not miss mixed infections, treats the patient at first visit, can be implemented at primary health care level, provides opportunity and time for education and counseling. The table below shows common causes of some STI syndromes.

Table 2.2: Identifying syndromes using syndromic approach

SYNDROME	**MOST COMMON CAUSE**
Vaginal discharge	Vaginitis(trichomniasis, candidisis) Cervicitis(gonorrhea, chlamydia)
Urethral discharge	Gonorrhea, chlamydia
Genital ulcer	Syphilis, chancroid, herpes
Lower abdominal pain	Gonorrhea, chlamydia, mixed anaerobes
Scrotal swelling	Gonorrhea, chlamydia
Inguinal bubo	LGV, Chancroid
Neonatal conjunctivitis	Gonorrhea, chlamydia

Findings from the current study strongly indicate the need for confirmatory research using biological testing for STIs in community samples of people living with HIV-AIDS. In the study by Liu et al, the syndromic management was compared with the "gold standard" of etiological tests, and the sensitivities and PPVs of WHO algorithms were 95% and 78% for urethral discharge syndrome, and 100% and 25% for genital ulcer syndrome in the males. These figures seemed to be lower than the results in a report from Shanghai and Chengdu in China [22]. Since the sample size for genital ulcer disease was small (55 in total), one might not draw sound conclusions from the study regarding the performance of syndromic management of genital ulcer disease.

Krishna Ray et.al (2008) [8] showed that in overal, self-reporting of morbidity was 65.0%. However, the percentage of women with some STD-related syndrome was 71.4%. The rural women were observed to have significantly more STD syndromes than their urban counterparts. The etiological diagnosis could be established in only 32.2% of cases. This study highlights the wide variation between self-reporting of morbidity and syndromic and etiology-based diagnosis in women from both rural and urban settings. This has implications for the syndromic approach to STI case management.

In a study by Desai et al. (2003), prevalence data was collected on 124 sex workers from the red light area of Surat, India. These women were mobilized to attend a health camp where they were given a clinical examination and specimens were collected for laboratory testing of STIs (syphilis, gonorrhea, Chlamydia, trichomoniasis, HIV and cervicitis). A behavioral interview was also conducted. Sensitivity, specificity and PPV was calculated to evaluate Indian syndromic management guidelines for VDS and GUs. They found that sensitivity for VDs to detect STIs was okay (60 to 80%), but that specificity was low (50 to 55%). PPV was very low (11% to 25%). Sensitivity of GUs to detect syphilis was low at 14.8% but specificity was high at 96.7%, the PPV was 57.1%. Authors concluded that syndromic management of STIs results in a high number of symptomatic cases going undetected and so alternative strategies for STI control need to be explored. Ching-Hui Tsai, MPH, Ta-Chung Lee et al. found out that for syndromic management, the sensitivity, specificity, and PPV detection of chlamydial, gonococcal, and combined forms of infection were 85.0%, 40.0% and 56.4%, respectively. In contrast, the sensitivity, specificity, and PPV for detection of syphilis were 78.8%, 18.1%, and 23.2%, respectively. Other studies that have looked at the comparison of syndromic diagnosis vs. etiological test are K Fonck, et al. (2000) and Wi, Mesola, et al. (1998) among others.

Chapter 3

Methodology

3.1 Study Design and Sampling Framework

A cross sectional multi site survey was conducted to determine prevalence and correlates of STI among HIV-infected adults in HIV programs throughout Kenya. All large HIV programs (39 clinics in 8 of 9 geographic regions serving 51% of all HIV-infected clients in care in Kenya) were visited by a mobile RTI screening team. Population-proportionate systematic sampling based on clinic population size was used.

Behavioral and clinical data, genital specimens, and blood were collected for testing. Specimens were mailed daily to central diagnostic laboratories for CD4 counts and HSV-2, trichomoniasis, gonorrhea, chlamydia, syphilis, bacterial vaginosis, and yeast testing.

3.2 Eligibility and Study Procedures

Participants were eligible to enroll in the study if they were HIV-infected and receiving care at one of the selected HIV care clinics, at least 18 years old, and able and willing to provide informed consent. women who were >36 weeks ges-

tation we excluded. The number of patients enrolled per site was proportionate to the number of registered patients receiving ART at that clinic. Depending on the daily recruitment goal and the total number of patients scheduled on a given day, the study team attempted to recruit and screen every n-th patient at the time of their routine clinic visit. If a patient was eligible and consented, a trained study nurse administered a structured questionnaire. In addition, the nurse assessed the patient's current STI symptoms, conducted a physical examination (including a pelvic examination for women), and collected specimens for laboratory testing. Study staff collected survey and clinic data on standardized paper forms and later entered these data directly into a customized web-based database. When applicable, the study nurse made a syndromic STI diagnosis and provided the participants with immediate treatment based on the current Kenya Ministry of Health guidelines on the management of STI. Once laboratory results were available, participants who were asymptomatic but diagnosed etiologically received appropriate treatment at a scheduled follow-up visit. Participants who received any STI treatment were given a referral card for their partners to seek STI care and HIV testing. The study was approved by the institutional review boards of the three collaborating institutions including the University of Washington, the U.S. Centers for Disease Control and Prevention, and the Kenya Medical Research Institute.

3.3 Specimens and Laboratory Testing

Selected STIs based on a combination of factors including expected prevalence, potential health benefits of detection and treatment, implications for HIV transmission, and the ability to conduct local timely testing were tested. Based on these criteria, tests for the following genital infections were conducted: chlamydia, gonorrhea, trichomoniasis, syphilis, herpes simplex virus type 2 (HSV-2), bacte-

rial vaginosis (women only), and vulvovaginal candidiasis (women only). Men provided a first catch urine specimen, while women had vaginal swabs collected during a pelvic exam. All participants had blood drawn to measure CD4 T-cell counts and for syphilis serology. Female participants also provided a urine sample for pregnancy testing.

3.4 Survey Measures

The nurse-administered survey included questions about socio-demographic information, recent sexual risk behaviors, and HIV/medical history. A subset of measures were chosen as potential correlates of STI. Specifically, participants provided data about their age, current marital status, and education level. Participants indicated the number of sexual partners they had in the past 3 months, and if they reported at least 1 partner, they were asked a series of questions about each of their most recent partners (up to 3). These questions included partner type, time since last sex, whether a condom was used during the last sexual encounter, and if either partner had been drinking during last sex. Participants were also asked to report if their clinic provided condoms. Finally, participants were asked when they were diagnosed with HIV, current ART use (and if so, when they began), current cotrimoxazole use, and circumcision status (among men only). All women were tested for pregnancy and recorded the results. Finally, each participant had their blood tested for CD4 count recorded cells/mm3.

In order to facilitate data analysis, some raw data variables required further manipulation and/or grouping resulting in the creation of new database variables. Refer to table 3.1 for details on coding of variable and variable types.

Table 3.1: Table of survey measures used in the analysis

Characteristic	Type	Categories
Response variables		
Any STI	Categorical	1 (positive), 0(negative)
Syndromic diagnosed STI	Categorical	1(positive),0(negative)
Independent variables		
Gender	categorical	1 (female) 0 (male)
Age	Categorical	1(18-24, 2 (25-34), 3 (35-44), 4 (45+)
Current marital status	Categorical	1(Single), 2(Monogamous), 3(marriage/cohabit), 4(Widowed), 5(Divorced/separated), 6(Polygamous marriage)
Education level	Categorical	1(Never gone school), 2(Primary and below), 3(Vocational/secondary), 4(Post-secondary and above)
Any sex partner	Categorical	1(Yes), 0(No)
Number of partners	Categorical	1(0), 2(1), 3(>1)
Condom use	Categorical	1 (Yes), 0 (No)
Alcohol drinking	Categorical	1(Yes), 0No)
Cotrimoxazole use	Categorical	1 (Yes), 0 (No)
Use of Haart	Categorical	1 (Yes), 0 (No)
Circumcised (men only)	Categorical	1 (Yes), 0 (No)
Pregnant (women only)	Categorical	1 (Yes), 0 (No)
CD4 count	categorical	1($\leq$ 250), 2(250-349), 3($\geq$ 350)

3.5 Statistical Analysis

Conventional descriptive statistics were used to assess the characteristics of the study participants. Prevalence was defined as the proportion of subjects with

positive tests at their visit. Covariates were considered in the analysis if they were associated with any STI in the literature. Model selection was done using Akaike Information Criteria. The model with the least value of AIC was chosen to be the best model. A multiple logistic regression model for a binomial response variable was fitted on the best model to obtain adjusted estimates of the odds ratios (AOR) and 95% confidence intervals (CI). Sensitivity and specificity and positive predictive value of syndromic diagnosis as compared to the lab diagnosis as the gold standard was calculated to validate syndromic diagnosis for syphilis,gonorrhea and trichomoniasis.

3.6 Theoretical overview of Logistic regression

3.6.1 Logistics Regression Model

The concern of logistic regression is on situations in which the outcome variable is dichotomous, although the theory has an extension to outcomes with three or more categories [7].The idea behind logistic regression modeling is closely related to multiple linear regression, except that the response variable is the logarithm of the odds of a certain outcome (of the response variable). Consider the case in which the response variable is binary. Then denote the response variable as y. Then, the random variable y would take values as 0 or 1, denoting failure or success of the required output. Suppose there are $(x_1, x_2, ..., x_n)$ explanatory variables that determine whether the outcome of$y$ would be 0 or 1. We wish to model the logarithm of the odds of the outcome $y = 1$, given the explanatory variables in the model. The specific form of the logistic regression model with unknown parameters $\beta = (\beta_0, \beta_1, \beta_2, ...\beta_n,)$ is

$$p(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2,...+\beta_n x_n}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2,...+\beta_n x_n}} \tag{3.1}$$

where $p(x)$ represents conditional probability $p(y = 1|x_1, x_2, ..., x_n)$

A transformation of $p(x)$ is called the logit transformation, and is given by:

$$\begin{aligned} \text{logit}p(x) &= ln\left\{\frac{p(x)}{1-p(x)}\right\} \\ &= ln\left\{\frac{p(y=1|x_1,x_2,...,x_n)}{1-p(y=1|x_1,x_2,...,x_n)}\right\} \\ &= \beta_0 + \beta_1 x_1 + \beta_1 x_1 + ... + \beta_n x_n \end{aligned} \tag{3.2}$$

The fraction

$$\frac{p(y=1|x_1,x_2,...,x_n)}{1-p(y=1|x_1,x_2,...,x_n)}$$

denotes the odds in favour of the outcome y= 1. The logarithm of the odds is taken in the model to enable implementation of an additive model (otherwise, probabilities/odds obey a multiplicative model). The last part of of the model (3.2) is a linear combination of the explanatory variables, with the log of odds taking the value β_0 if the explanatory variables are all zero (not influential).

3.6.2 Maximum Likelihood (ML) Estimation of the Parameters

Suppose we have a sample of n independent observations (y_i, x_i) i= 1,2,...n , where y_i denotes the value of a dichotomous outcome variable, and x_i is the value of the explanatory variables for the i-th subject. Assume $y \sim Bernoulli\{1, p(x)\}$ $i = 1, 2...n$

Based on a set of data, we estimate the parameter vector $\beta = (\beta_0, \beta_1, \beta_2, ...\beta_n,)$ to fit the logistic regression model in equation (3.3). To find the ML estimator of β , we define the likelihood function as follows

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} p(x_i)^{y_i}(1-p(x_i))^{1-y_i} \qquad (3.3)\\
&= \prod_{i=1}^{n} \left[\frac{\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}}{1-\frac{e^{x_i'\beta}}{1-e^{x_i'\beta}}}\right]\left[\frac{1+e^{x_i'\beta}-e^{x_i'\beta}}{1+e^{x_i'\beta}}\right]\\
&= \prod_{i=1}^{n} \left[\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \times \frac{1+e^{x_i'\beta}}{1}\right]\left[\frac{1}{1+e^{x_i'\beta}}\right]\\
&= \prod_{i=1}^{n} \left[\frac{(e^{x_i'\beta})_i^y}{1+e^{x_i'\beta}}\right]
\end{aligned}
$$

Now, we find the ML estimates, $\hat{\beta}$, of β by maximizing the log-likelihood function for the observed values of y_i and x_i . Since maximizing the log of a function is equivalent to maximizing the function, we often work with the log-likelihood because it is generally less cumbersome to use for mathematical operations, such as differentiation. Therefore, the loglikelihood function yields,

$$
l(\beta) = \sum_{i=1}^{n} y_i x_i'\beta - \sum_{i=1}^{n} \log(1+e^{x_i'\beta}) \qquad (3.4)
$$

The first derivative of the log-likelihood function gives the gradient.

We have the first derivative of $x_i'\beta$ with respect to β_j is x_{ij} , so

$$
\begin{aligned}
\frac{\delta l(\beta)}{\delta \beta_j} &= \sum_{i=1}^{n} y_i x_{ij} - \sum_{i=1}^{n} \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\\
&= \sum_{i=1}^{n} y_i x_i'\beta - \sum_{i=1}^{n} p(x_i)x_{ij}\\
&= \sum_{i=1}^{n} (y_i - p(x_i))x_{ij} \qquad (3.5)
\end{aligned}
$$

The second derivatives are

$$\begin{aligned}
\frac{\delta^2 l(\beta)}{\delta\beta_j \delta\beta_k} &= -\sum_{i=1}^{n} x_{ij} \frac{\delta}{\delta\beta_k}\left(\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right) \\
&= -\sum_{i=1}^{n} x_{ij}\left[\frac{(1+e^{x_i'\beta)} \times e^{x_i'\beta} x_{ij} - e^{x_i'\beta} \times e^{x_i'\beta} \times x_{ij}}{1+e^{x_i'\beta}}\right] \\
&= -\sum_{i=1}^{n} x_{ij}x_{ik}\left[\frac{e^{x_i'\beta}(1+e^{x_i'\beta} - e^{x_i'\beta})}{1+e^{x_i'\beta}}\right] \\
&= -\sum_{i=1}^{n} x_{ij}x_{ik}p(x_i)(1-p(x_i))
\end{aligned} \tag{3.6}$$

The maximum likelihood estimates for β can be found by setting each of the j + 1 equations in (3.5) equal to 0 and solving for each β_j

Each such solution, if any exists, specifies a critical point-either a maximum or a minimum.

The critical point will be a maximum if the matrix of second partial derivatives is negative definite; that is, if every element on the diagonal of the matrix is less than zero Another useful property of this matrix is that it forms the variance-covariance matrix of the parameter estimates. It is formed by differentiating each of the j + 1 equations in (3.5) a second time with respect to each element of β, denoted by β_k . The general form of the matrix of second partial derivatives is given by equation (3.6).

3.6.3 The Newton-Raphson Method

Setting the equations in (3.5) equal to zero results in a system of j + 1 nonlinear equations each with j + 1 unknown variables. The solution to the system is a vector with elements, j. After verifying that the matrix of second partial derivatives is negative definite, and that the solution is the global maximum rather than a local maximum, then we can conclude that this vector contains the parameter estimates for which the observed data would have the highest probability of occurrence. However, solving a system of nonlinear equations is not easy—the

solution cannot be derived algebraically as it can in the case of linear equations. The solution must be numerically estimated using an iterative process. Perhaps the most popular method for solving systems of nonlinear equations is Newton's method, also called the Newton-Raphson method. Newton's method begins with an initial guess for the solution then uses the first two terms of the Taylor polynomial evaluated at the initial guess to come up with another estimate that is closer to the solution. This process continues until it converges (hopefully) to the actual solution.

The Newton-Raphson algorithm requires the second-derivatives

$$\frac{\delta^2 l(\beta)}{\delta\beta_j \delta\beta_k} = -\sum_{i=1}^{n} x_{ij} x_{ik} p(x_i)(1 - p(x_i))$$

Starting with β^{old} , a single Newton-Raphson update is

$$\beta^{new} = \beta^{old} - (\frac{\delta^2 l(\beta)}{\delta\beta_j \delta\beta_k})^{-1} \frac{\delta l(\beta)}{\delta\beta_j} \tag{3.7}$$

where the derivatives are evaluated at β^{old}.

3.6.4 Odds and Odds Ratio

The odds ratio is a measure of association, which quantifies the relationship between an exposure and health outcome from a comparative study. It is the ratio of the odds in favor of getting the disease, if exposed, to the odds in favor of getting the disease, if not exposed. Cox (1970) discussed some general advantages of the odds ratio as a measure of association for binary responses. Bland and Douglas (2000) mentioned that there are mainly three reasons to use the odds ratio. Firstly, they provide an estimate (with confidence interval) for the relationship between two binary variables. Secondly, they enable us to examine the effects of other variables on that relationship, using logistic regression. Thirdly, they have a special and very convenient interpretation. Therefore, it is essential to introduce the terms odds and odds ratio in order to discuss binary data and to

interpret the logistic regression coefficients. For a probability of success,the odds are defined to be

$$odd = \frac{p(x)}{1 - p(x)}$$

The odds are nonnegative, with odds >1.0 when a success is more likely than a failure.

In a 2 by 2 table, the probability of success is $p_1(x)$ in row 1 and $p_2(x)$ in row 2. Within row 1, the odds of success are defined to be

$$odd_1 = \frac{p_1(x)}{1 - p_1(x)}$$

and within row 2, the odds of success are defined to be

$$odd_2 = \frac{p_2(x)}{1 - p_2(x)}$$

The ratio of odds from the two rows is called the odds ratio, which is given by

$$odds\ ratio = \frac{\frac{p_1(x)}{1-p_1(x)}}{\frac{p_2(x)}{1-p_2(x)}}$$

3.6.5 Model Selection

In case of a large number of predictors, it is often desirable to determine a smaller subset with the strongest effects. Our first strategy was to consider stepwise selection with Akaike's Information Criterion (AIC), as defined by

$$AIC = -2logL(M) + 2 * K$$

Here, L(β) is the likelihood function and k is the number of parameters included in the model. The basic idea behind the information criteria is penalizing the likelihood for the model complexity - the number of explanatory variables used in the model. With this method, model with the least value of AIC is the best model.

There exists other approaches that can be applied but were not of the interest to this project. These include:

Mallow's CP

It selects the model that minimizes the mathematical expectation of scaled sum of squared error , where the error is defined as the algebraic distance between the predicted and observed data.

Minimum Description Length (MDL)

Was introduced by Rissanen. The underlying logic of MDL is that the simplest model that sufficiently describes the data is the best model.

Bayesian Information Criterion (BIC)

BIC chooses the model that maximizes the conditional probability of describing a data set by a model constrained by some priori information.

3.6.6 Other Logistic Regression Applications

There are many logistic regression models that are not of the standard form as given earlier. The following are some of the models that are often used by different statisticians.

Conditional Logistic Regression

Conditional logistic regression is useful in investigating the relationship between an outcome and a set of prognostic factors in a matched case-control studies, the outcome being whether the subject is a case or a control.

Bradley-Terry Model for Paired Comparison

The Bradley-Terry Model is useful in establishing the overall ranking of n items through paired comparisons. For instance, it is difficult for a panelist to rate all 9

brands of beer at the same occasion; rather it is preferable to compare the brands in a pair wise manner. For a given pair of products, the panelist would state his preference after tasting them at the same occasion.

Multinomial Logistic Model

This is a regression model which generalizes logistic regression by allowing more than two discrete outcomes. The multinomial logit model is useful in investigating consumer choice behavior and has become increasingly popular in marketing research.

3.7 Sensitivity, Specificity and Positive Predictive Value

The words "sensitivity" and "specificity" have their origins in screening tests for diseases. When a single test is performed, the person may in fact have the disease or the person may be disease free. The test result may be positive, indicating the presence of disease, or the test result may be negative, indicating the absence of the disease. The table below displays test results in the columns and true status of the person being tested in the rows.

		test results	
		Positive(+)	negative(-)
True Status of Nature(S)	Disease(+)	a	b
	No Disease (-)	c	d

Table 3.2: Calculating sensitivity and specificity

Sensitivity:We define sensitivity as the probability that the test says a person has the disease when in fact they do have the disease.

This is

$$P(\frac{T^+}{S^+}) = \frac{a}{a+b} \tag{3.8}$$

Sensitivity is a measure of how likely it is for a test to pick up the presence of a disease in a person who has it.

Specificity:We define specificity as the probability that the test says a person does not have the disease when in fact they are disease free.

This is

$$P(\frac{T^-}{S^-}) = \frac{d}{c+d} \tag{3.9}$$

Positive predictive value:We define positive predictive value as proration of patients who test positive who actually have the disease.

This is

$$P(\frac{S^+}{T^+}) = \frac{a}{a+c} \tag{3.10}$$

Chapter 4

Results and Analysis

4.1 Population Characteristics

Of the 1661 participants enrolled in the study, table 4.1, 64% (1063) were female and 36% (598) were female. Majority of the participants were aged between 35-44 years old (38.01%) followed by those between age 25-34 (32.39%). Those of age 45 and above were 24.71% and the least group were those in age bracket of 18-24 (4.89%). More than half of the participants were either married (monogamous marriage) or cohabiting with their partners (55.19%). 13.63% were single, 15.20% were widowed, 10.13% were either divorced or separated. Only 5.85% were in a polygamous marriage. More than a half of the participants have attended at least a primary education (53.99%) followed by vocational training or secondary education at 38.47% and lastly post-secondary education and above at 54%.

40.24% of participants reported no sexual partners in the past 3 months. Among those who did have a recent partner, the vast majority reported only 1 partner in the past 3 months. Over half of all participants stated that they last had sex in the past week; 70.65% of reported using a condom during this encounter. Only 78 (8.07%) had sex under the influence of alcohol.

In table 4.2, among all participants, approximately 75% were on ART, with just

Table 4.1: Descriptive statistics for participant's socio-demographic and sexual characteristics (N=1,661)

Characteristic	N	%
Gender		
Female	1063	64
Male	598	36
Age		
18-24	81	4.89
25-34	536	32.39
35-44	629	38.01
45+	409	24.71
Current marital status		
Single	226	13.63
Monogamous marriage/cohabiting	915	55.19
Widowed	252	15.20
Divorced/separated	168	10.13
Polygamous marriage	97	5.85
Education level		
Primary school and below	845	53.99
Vocational/secondary	602	38.47
Post-secondary and above	118	7.54
Sexual behavior		
Number of partners in past 3 months		
0	664	40.24
1	934	56.61
> 1	52	3.15
Condom use at last sex	691	70.65
Alcohol drinking during sex (either partner)	78	8.07

under 50% having $\geq 350\, cells/mm^3$ and approximately one-third with $\leq 250 cells/mm^3$. Sixty percent of men were circumcised, and 4% of women were pregnant.

Table 4.2: Descriptive statistics for Participants HIV/medical characteristics (N=1,661)

Characteristic	**N**	**%**
Cotrimoxazole use	1,023	64.18
Use of Haart	1,226	75.35
Circumcised (men only)	338	59.72
Pregnant (women only)	38	3.80
CD4 count (cells/mm3)		
≤ 250	510	31.27
250-349	342	20.97
≥ 350	779	47.76

4.2 Prevalence of STI and Other Genital Infections

Figure 4.1 illustrate the prevalence STI and other genital infections. 9.81% of HIV-infected participant had an etiologic diagnosis with majority being trichomoniasis (8.04% overall prevalence). There was much lower prevalence of gonorrhea (1.33%), chlamydia (0.36%), and syphilis (0.48%). Sixty nine percent of the participants had positive serology tests for HSV-2. Among women who participated in the study, 18.6% were diagnosed with bacterial vaginosis.

Figure 4.2 illustrate the distribution of syndromic diagnosis. Those who had any symptoms of STI were found to be 21.73%. This composed of either virginal discharge (25.02%) or PID (7.45%) among women and urethral discharge (4.67%) among men. 4.34% of all participate were infected with genital ulcer.

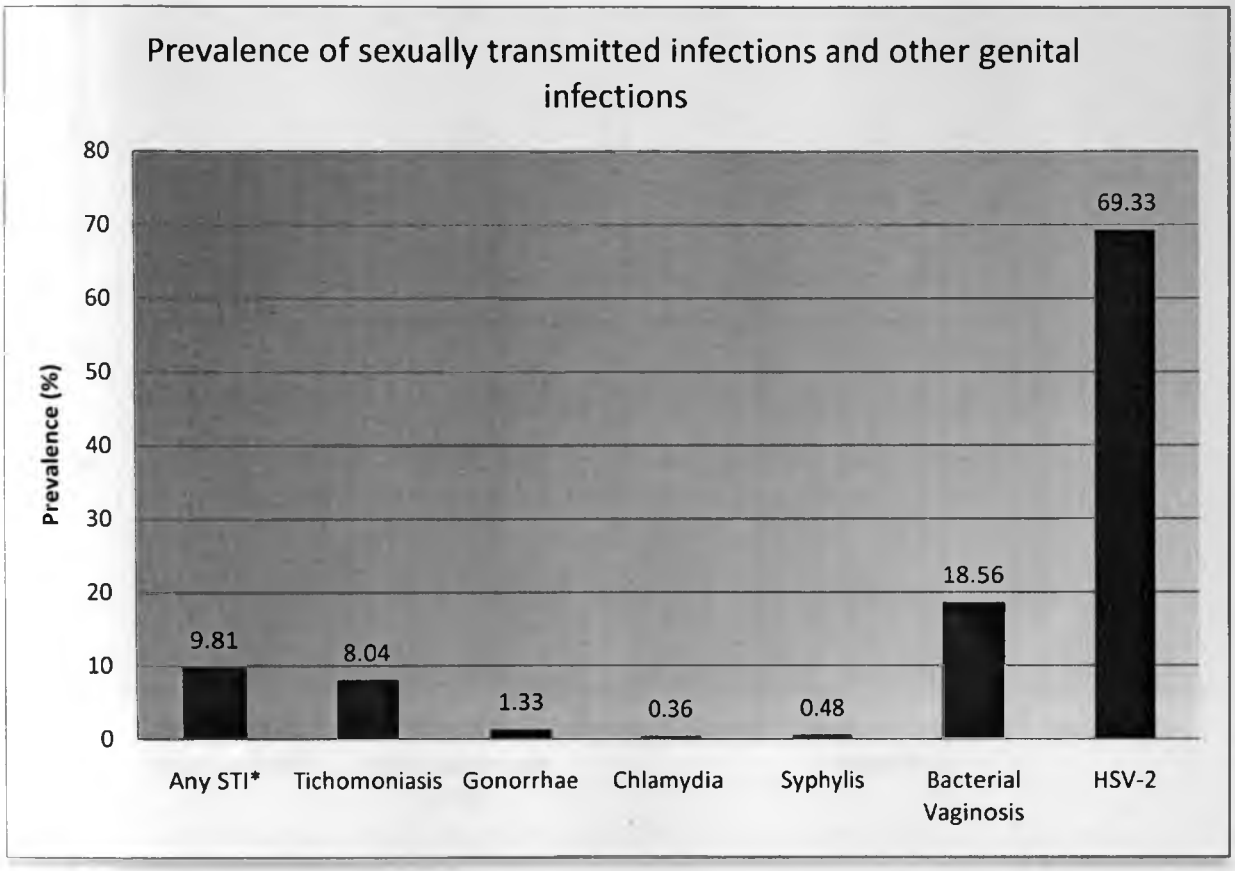

Figure 4.1: Prevalence of STI and Other Genital Infections

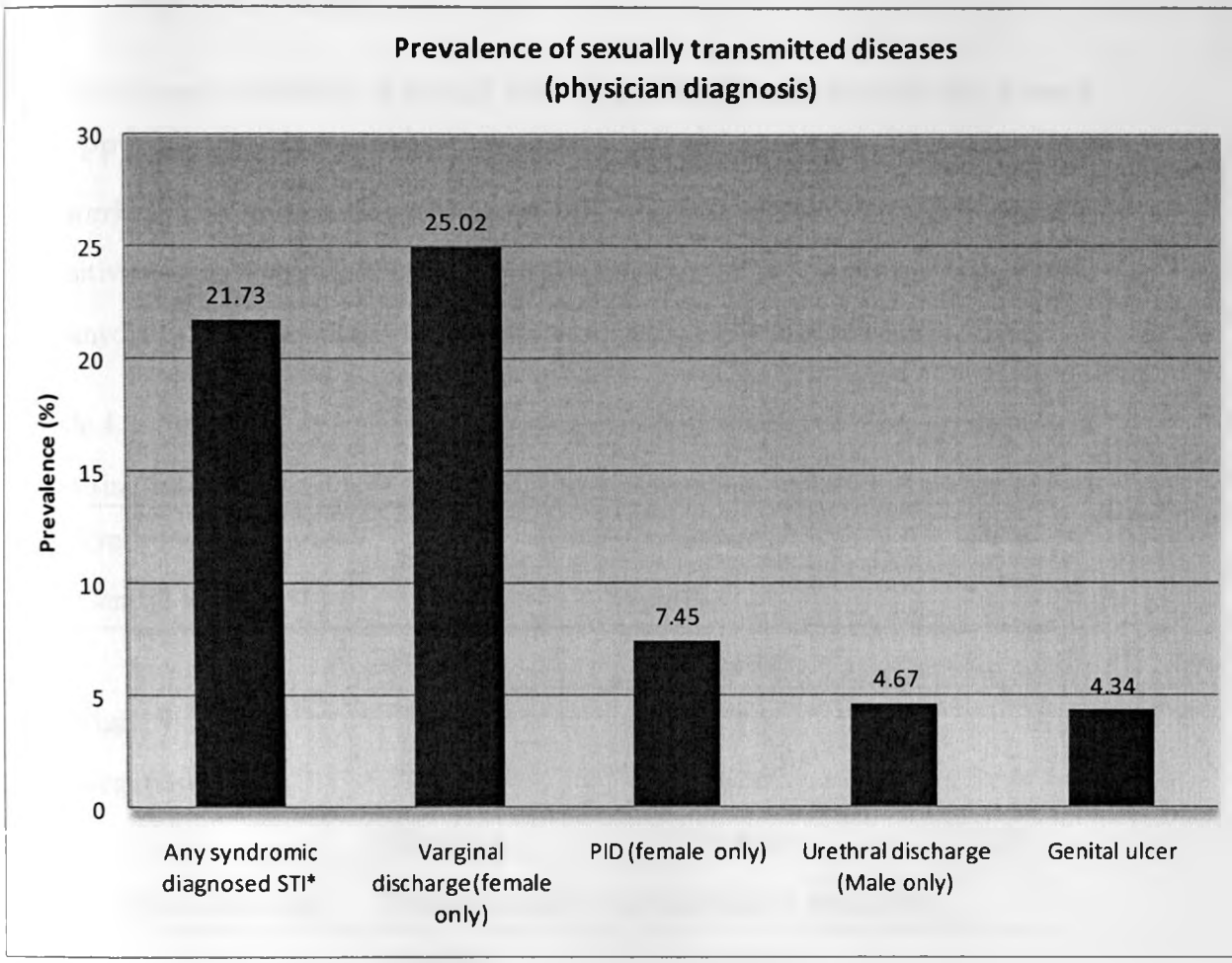

Figure 4.2: Prevalence of STI Syndromes

4.3 Syndromic Diagnosis vs Etiological Diagnosis of STDs

Sensitivity and specificity of genital ulcer in predicting syphilis were zero percent and PPV was also low at 0.00%. Sensitivity of urethral discharge for predicting Gonorrhea or chlamydia was 20%, specificity was high at 95% while PPV was 4%. Sensitivity, specificity and PPV of virginal discharge for predicting Gonorrhea, chlamydia or trichomoniasis in females were 36%, 76%, and 18% respectively.

Table 4.3: Sensitivit, specificity and positive predictive value of three syndromes in detecting lab confirmed STI (Syphilis, Gonorrhea,chlamydia and Trichomoniasis)

Syndromic diagnosis	Etiologic diagnosis		PPV
Genital ulcer	Syphilis		
	Positive	Negative	
Positive	0	72	
Negative	8	1574	
	SE=0.00	SP=0.96	0.00
Varginal discharge	Trichomoniasis or chlamydia or gonorrhea		
	Positive	Negative	
Positive	47	219	
Negative	85	712	
	SE=0.36	SP=0.76	0.18
Urethral discharge	Chlamydia or gonorrhea		
	Positive	Negative	
Positive	1	27	
Negative	4	567	
	SE=0.20	SP=0.95	0.04

Summarizing table 4.3 we have

Table 4.4: Summary table

Syndromic diagnosis

	Sensitivity	Specificity	Positive predictive value
Genital Ulcer	0.00	0.96	0.00
Varginal discharge	0.36	0.76	0.18
Urethral discharge	0.20	0.95	0.04

4.4 Correlates of STI

4.4.1 Bivariate Analysis of Correlates of STI

In bivariate analysis, we explored socio-demographic, sexual behavior, and health-related correlates of any non-viral STI diagnosis, defined as laboratory-confirmed trichomoniasis, gonorrhea, chlamydia, or syphilis . In table 4.5(socio-demographics characteristics), age, gender and education level, were found to be associated with prevalent STI (all with P value < 0.001). However , there was no significant difference in the marital status in STI prevalence. in table 4.6, use of alcohol was also found to be associated with the STI prevalence (P=0.043). Number of sexual partners was slightly significant at p value =0.067. On the medical characteristics, only the circumcision in men was found to be associated with the STI prevalence (p=0.039).

Since not all the variables were significatively associated with any STI, there was need to choose the best model with variables that best explain the outcome. A model selection criteria was applied to get the model. The results were as shown in table 4.7.

Table 4.5: Bivariate analysis of covariates (socio-demographic characteristics) of any STI (Chi square test)

	Negative		Positive		P value
Socio-demographic	No.	%	No.	%	
Gender					
Female	928	61.9	135	82.8	
Male	570	38.1	28	17.2	< 0.001
Age of Participant					
18-24	67	4.5	14	8.6	
25-33	469	31.4	67	41.1	
34-44	565	37.9	64	39.3	< 0.001
45+	391	26.2	18	11	
Current Martial Status					
Single	198	13.2	28	17.2	
Married/Cohabiting	837	56	78	47.9	
Widow	228	15.3	24	14.7	0.221
Divorced/Sep	146	9.8	22	13.5	
Polygamous Marriage	86	5.8 11	6.7		
Highest Education					
Primary or lower	736	49.1	109	66.9	
Secondary/Vocational	558	37.2	44	27.0	< 0.001
College/University	117	7.8	1 0.6		
No school	87	5.8	9	5.5	

4.4.2 Model Selection

The analysis for this study involved logistic regression modeling and model comparison. All models used any STI as the dependent variable. Akaike information criteria (AIC) was used as a model selection criteria to choose the best model. Table 4.7 describes the candidate models and AIC values for model comparizon. Full model (Model 1) contained the following variables: gender , age , education condom, alcohol , sex partner , CD4 count, Number of sex partners' marital status , CTX, HAART, Pregnancy , circumcision. A model containing gender, age, education, and Number of sex partners (model 10) was found to be the best model(AIC=1004.55).

4.4.3 Multiple Regression Analysis Results

Table 4.8 shows the results of multiple logistics regression. After adjusting for factors that were included in the best model, all the four variables were significantly associated with the presence of any STI. The odds of having STI for male was 0.33 (95% CI : 0.20- 0.53) compared to women.This means that being a man is 76% less like to get any STI than being a woman. The odds of having STI for the participants aged 45 years and above was found to be 0.33 (95% CI : 0.15 - 0 .71) as compared to those of age between18 to 24 years old. In addition, those with vocational training/secondary education and post secondary education and above had odds ratio of 0.56 (95% CI : 0.39- 0.82) and 0.05 (95% CI : 0.01-0.38). Those with more than one sexual partner was significantly associated with having STI, odds ration of 3.45 (95% CI: 1.43-8.30). This means that those with more than one sexual partner were 3.5 times more likely to have STI compared to those with no sexual partners.

Table 4.6: Bivariate analysis of covariates(Sexual behavior and medical characteristic) of any STI (Chi square test)

	Negative		Positive		P value
Sexual behavior					
Number of Partners in Past 3 Months					
0	611	41.1	53	32.7	
1	833	56	101	62.3	0.067
>1	44	3	8	4.9	
Condom at Last Sex					
No	253	29.1	34	31.2	
Yes	616	70.9	75	68.8	0.653
Alcohol at Last Sex					
No	796	92.6	93	86.9	
Yes	64	7.4	14	13.1	0.043
Medical characteristics					
Pregnant (Women Only)					
No	841	96.1	120	96.8	
Yes	34	3.9	4	3.2	0.719
Circumcised (Men Only)					
No	212	39.3	16	59.3	
Yes	327	60.7	11	40.7	0.039
HAART Use					
No	354	24.1	47	29.6	
Yes	1,114	75.9	112	70.4	0.13
Cotrimoxazole Use					
No	514	35.8	57	35.8	
Yes	921	64.2	102	64.2	0.994
CD4 Count					
< 250	451	30.6	59	37.3	
250-349	309	21	33	20.9	0.185
350+	713	48.4	66	41.8	

Table 4.7: Stepwisw AIC process for getting the best model

Model	Variable in the model	Number of parameters	Value of AIC
1	model1	13	1027.45
2	model1- Type of sex partner	12	1027.45
3	model2-marital status	11	1022.59
4	model4-condom use	10	1019.05
5	model5-HAART use	9	1012.23
6	model6- CTX use	8	1009.03
7	model7- Alcoho use	7	1019.05
8	model8- Pregnancy	6	1006.49
9	model9-circumcision	5	1004.88
10	model1-CD4 count	4	1004.55

Table 4.8: Results of multiple logistics regression

Characteristics	adjORb	95% CI
Gender		
male	0.33	**0.20- 0.53**
Female	1.00	
Age		
18-24	1.00	
25-34	0.75	0.39 - 1.44
35-44	0.67	0.35 - 1.27
45+	0.33	**0.15 - 0 .71**
Education level		
Never attended school	0.86	0.41 - 1.82
Primary school and below	1.00	
Vocational/secondary	0.56	**0.39- 0.82**
Post-secondary and above	0.05	**0.01-0.38**
Number of partners (past 3 months)		
0	1.00	
1	1.43	0.97-2.11
< 1	3.45	**1.43-8.30**

Chapter 5

Discussion

Prevalence of any STI (lab confirmed chlamydia, syphilis, gonorrhea and trichomoniasis) was relatively low compared to results by kalicham and colligues whose results yielded prevalence of any STI at 16.8%. Among these STI, trichomoniasis was overwhelmingly the most common genital non-viral infection, while HSV-2 had the highest overall prevalence (>65%). This is consistent with a study conducted among HIV infected pregnant women in Kenya. We found a lower prevalence of syphilis than among HIV-infected adults in the 2007 KAIS. In contrast to this study, the vast majority of HIV-infected KAIS respondents were unaware of their HIV status. Because participants in this study were all enrolled in HIV care programs, they may be individuals with less risky sexual behavior or be more likely to receive STI treatment for symptomatic STI.

To our knowledge, there have been only two other large STI surveillance studies among HIV-infected adults in Africa, neither in HIV-1 care programs. A 1993-1997 study of male factory workers in Harare, Zimbabwe, found that HIV-positive men had an incidence of self-reported STI syndromes (including urethral discharge, genital ulcer, genital warts) of 16.8 per 100 person-years [12]. Because of the highly selected population and the syndromic, self-report approach to iden-

tifying STI, it is difficult to compare these findings to ours. The other African study enrolled HIV-infected pregnant women in Malawi, Tanzania, and Zambia between 2001-2003, and reported prevalence for bacterial vaginosis (47.8%), vaginal candidiasis (22.4%), trichomoniasis (18.8%), chlamydia (2.6%), genital ulcers (2.2%), and gonorrhea (1.7%)14. These prevalence are all higher than what we observed in our study, which may be attributable to regional differences.

We found that being a female, younger age, less education, and more recent sexual partners were all associated with an increased risk of STI, consistent with other African studies[5]. A higher CD4 cell count ($\geq$ 350 cells/mm3) was associated with reduced risk of STI but only in the bivariate analysis. CD4 count variable was not so important during model selection process and was not included in the logistics regression analysis. Among men, circumcision status was significantly associated with STI in bivariate analysis. Data from a meta-analysis by Weiss et al. suggested a protective effect of male circumcision on syphilis, and to a lesser extent, HSV-2. Use of alcohol during sex was also found to be associated with any STI in the bivariate analysis. Only few studies supported this as was seen in the literature.

The results indicated that urethral discharge could only identify 20% of true cases of gonorrhea or chlamydia in male. On the other hand virginal discharge could identify 47% of those whose suffered from either gonorrhea, chlamydia or trichomoniasis in female participants. It seems that a small proportion of patients are treated when they are actually suffering from the disease.

Sexually transmitted genital ulcer diseases are mainly caused by herpes and syphilis. We found that none of the true cases of syphilis could be identified. The studies conducted in Africa and South America indicated that the ability to detect syphilis using a syndromic approach was good, with sensitivities ranging

from 72%-100%.(8,10,20,21). This so off the results in this study. The data confirmed that syndromic diagnosis of syphilis is unsatisfactory. This might be due to variation in clinical manifestations and the asymptomatic nature of syphilis. The qualification of the clinicians who were carrying out the syndromic diagnosis to identify the syndrioms could be an issue.

5.1 Study Limitations

Although the sample of data used in the analysis is representative of HIV-infected adults enrolled in HIV care clinics throughout Kenya, the 2007 KAIS data demonstrated that many HIV-infected Kenyans were unaware of their status [17], thus we cannot say that this study is representative of all HIV-infected adults in Kenya.

The sexual behavior measures we assessed may have been subject to social desirability bias if participants who engaged in high risk sexual behaviors were less likely to report this.

Due to the fact that the binary regression models are nonlinear, no single approach to interpretation can fully describe the relationship between a variable and the outcome. In general, the estimated parameters from the binary regression model to provide useful information for understanding the relationship between independent variable and the outcome. With the exception of the rarely used method of interesting the latent variable,substantive meaningful interpretations are based in the predicted probabilities and functions of those probabilities [11].

Studies that were used to compare results from this study were different in terms of the population composition. This study looked at both genders while the other were looking at either only male or female and were not necessarily HIV infected.

Chapter 6

Conclusion and Recommendation

Due to inconsistent results in STI prevalence, probably due to difference in geographical regions, more studies among HIV infected individuals need to be conducted in Africa, especially in Kenya to get true picture of the situation. This will help in curbing the spread of STI not only to HIV infected individuals but to the entire population.

Community sensitization on the risk factors of STI should be conducted to help reduce the spread of STI.

Syndromic management of urethral discharge, virginal discharge and Genital ulcer were insensitive, while specificity were relatively good but not satisfactory. This study recommends further validation of syndromic diagnosis approaches for virginal discharge,urethral ulcer and genital ulcer. Also alternative strategies for STI control need to be explored.

Appendix A

R codes for data analysis

```
   Importing data in the R
####################################
sti4=read.csv("sti4.csv")
sti
attach(sti4)
names(sti4)

Changing the responses to factors
#######################################################
sti4$pregnant= as.factor(sti4$pregnant)
sti4$circum= as.factor(sti4$circum)
sti4$anysex31= as.factor(sti4$anysex31)
sti4$sexpart31= as.factor(sti4$sexpart31)
sti4$agecat1= as.factor(sti4$agecat1)
sti4$haart= as.factor(sti4$haart)
sti4$circum= as.factor(sti4$circum)
sti4$cotrim= as.factor(sti4$cotrim)
sti4$condomlast1= as.factor(sti4$condomlast1)
```

```
sti4$edu= as.factor(sti4$edu1)
sti4$marital2= as.factor(sti4$marital2)
sti4$alclast1= as.factor(sti4$alclast1)
sti4$cd4cat= as.factor(sti4$cd4cat)
sti4$anysynddiag= as.factor(sti4$anysynddiag)
sti4$anystidiag= as.factor(sti4$anystidiag)

#labelling variables
##############################################################################

edu=factor(levels = c(1,2,3),labels = c("Primary or lower", "Secondary/Vocat
"College/University"))
circum=factor(levels = c(1,0),labels = c("Yes", "No"))
anysex3=factor(levels = c(1,0),labels = c("Yes", "No"))
haart=factor(levels = c(1,0),labels = c("Yes", "No"))
cotrimt=factor(levels = c(1,0),labels = c("Yes", "No"))
condomlast=factor(levels = c(1,0),labels = c("Yes", "No"))
alclast=factor(levels = c(1,0),labels = c("Yes", "No"))
cd4cat=factor(levels = c(1,2,3),labels = c("<250","250-349","350+"))
marital=factor(levels = c(1,2,3,4,5),labels = c("Single","Married/Cohab",
"Widow","Divorced/Sep","Polygamous Marriage"))
agecat=factor(levels = c(1,2,3,4),labels = c("18-24","25-34","35-44","45+"))
anysynddiag=factor(levels = c(1,0),labels = c("Yes", "No"))
anystidiag=factor(levels = c(1,0),labels = c("Yes", "No"))

#Frequency table (Univariate analysis)
#####################################################################
Table1=table(pregnant)
```

```
round(100*.Table/sum(.Table1), 2)
Table2=table(circum)
round(100*.Table/sum(.Table2), 2)
Table3=table(anysex31)
round(100*.Table/sum(.Table3), 2)
Table4=table(sexpart31)
round(100*.Table/sum(.Table4), 2))
Table5=table(agecat1)
round(100*.Table/sum(.Table5), 2)
Table6=table(haart)
round(100*.Table/sum(.Table6), 2)
Table7=table(cotrim)
round(100*.Table/sum(.Table7), 2)
Table8=table(condomlast)
round(100*.Table/sum(.Table8), 2)
Table9=table(edu1)
round(100*.Table/sum(.Table9), 2)
Table10=table(marital2)
round(100*.Table/sum(.Table10), 2)
Table11=table(alclast)
round(100*.Table/sum(.Table11), 2)
Table12=table(cd4cat)
round(100*.Table/sum(.Table12), 2)
Table13=table(anysynddiag)
round(100*.Table/sum(.Table13), 2)
Table14=table(anystidiag)
round(100*.Table/sum(.Table14), 2)
```

```
Bivariate analysis using chi square test
##################################################
chisq.test(table(anystidiag,anysex3))
chisq.test(table(anystidiag,condomlast1))
chisq.test(table(anystidiag,cd4cat))
chisq.test(table(anystidiag,agecat1))
chisq.test(table(anystidiag,marital2))
chisq.test(table(anystidiag,cotrim))
chisq.test(table(anystidiag,circum))
chisq.test(table(anystidiag,edu1))
chisq.test(table(gender,anystidiag))
chisq.test(table(anystidiag,sexpart31))
chisq.test(table(anystidiag,haart))
chisq.test(table(anystidiag,cd4cat))
chisq.test(table(anystidiag,edu1))
chisq.test(table(anystidiag,alclast1))


library(MASS)
Model selection (AIC)
#########################################################################
GLM.1 = glm(anystidiag ~ gender + agecat1 +edu1 +condomlast1 + alclast1
+anysex31 + cd4cat +sexpart31 + marital2+ cotrim+haart+
pregnant+circum, family=binomial(logit), data=sti4)
summary(GLM.1)
stepAIC(GLM.1)
```

```
Logistics regression analysis
##############################################################################
lin = glm(formula = anystidiag ~ gender + agecat1 + edu1 + sexpart31,
family = binomial(), data = sti4)
summary(lin)
exp(coef(lin))

#For sensitivity and specificity - Univeriate analysis
#############################################################3

table(gudiag syphdiag)
table(vddiag femaled)
table(uddiag maled)

rm(list=ls(all=TRUE))  # for clearing data from memory
```

Bibliography

[1] Bland J. Martin and Altman G. Douglas (2000). Statistics Notes: The odds ratio. *British Medical Journal*, 320,1468.

[2] Ching-Hui Tsai , et al:The cost-effectiveness of syndromic management for male STD patients with urethral discharge symptoms and genital ulcer disease in Taiwan.

[3] Cox, D. R. (1970). Analysis of Binary Data. London: *Chapman and Hall*

[4] Desai, V. K., J. K. Kosambiya, et al. (2003). "Prevalence of sexually transmitted infections and performance of STI syndromes against aetiological diagnosis, in female sex workers of red light area in Surat, India." *Sex Transm Infect* 79(2): 111-115.

[5] Ed. Lester Breslow. Gale Cengage, 2002: "Sexually Transmitted Diseases."emphEncyclopedia of Public Health Encyclopedia of Public Health **eNotes.com**. 2006. 9 Jun, 2011 *http://www.enotes.com/public-health-encyclopedia/sexually-transmitted-diseases*

[6] K Fonck, N Kidula, W Jaoko, B Estambale, P Claeys, J Ndinya-Achola, P Kirui, J Bwayo, and M Temmerman: Validity of the vagi-

nal discharge algorithm among pregnant and non-pregnant women in Nairobi, Kenya.*Sex Transm Infect.* 2000 Feb;76(1):33-8

[7] Hosmer, D. and Lemeshow, S. (1989). Applied Logistic Regression. *John Wiley and Sons Inc.*, New York.

[8] Krishna Ray, Sumathi Muralidhar, Manju Bala, Madhu Kumari, Sudha Salhan, S.M. Gupta, M. Bhattacharya :Comparative study of syndromic and etiological diagnosis of reproductive tract infections/sexually transmitted infections in women in Delhi. *International Journal of Infectious Diseases - November 2009* (Vol. 13, Issue 6, Pages e352-e359, DOI: 10.1016/j.ijid.2008.11.021)

[9] Lavreys L, Rakwar JP, Thompson ML, et al. Effect of circumcision on incidence of Human Immunodeficiency Virus Type 1 and other sexually transmitted diseases: a prospective cohort study of trucking company employees in Kenya. *J Infect Dis* 1999, 180: 330-336

[10] Liao, Tim F. 1994. Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models. *Thousand Oaks: SAGE Publication*

[11] Long J.S; Freese J. Rgression models for categorical dependent variables using STATA, pages 131-134.

[12] Machekano et.al. Report of sexually diseases by HIV infected men during follow up:time to target the HIV infected? *Sex Trans Infect.*June 2000;76(3):188-192.

[13] McClelland RS, Lavreys L, Katingima C, et al. Contribution of HIV-1 infection to acquisition of sexually transmitted disease: a 10-year prospective study.*J Infect Dis* 2005;191:333-8.

[14] Merson MH. Slowing the spread of HIV: agenda for the 1990s. Science 1993;220:1266-1268)

[15] Mehta.CR and Patel N.R(1995). Exact Logistics regression:Therory and examples, statistics in Medicine,14,pp 2143-216

[16] Moses S, Bailey RC, Ronald AR. Male circumcision: assessment of health benefits and risks.*Sex Transm Inf* 1998, 74: 368-373

[17] National AIDS and STI Control Programme, Ministry of Health, Kenya. July 2008. *Kenya AIDS Indicator Survey 2007*: Preliminary Report. Nairobi, Kenya.

[18] Sabrina S. Plitt, phd, Richard S. Garfein, phd, Charlotte A. Gaydos, drph, steffanie A. Strathdee, phd, and Susan G. Sherman, phd, and Tha E. Taha, phd. Prevalence and Correlates of Chlamydia trachomatis, Neisseria gonorrhoeae, Trichomonas vaginalis infections, and Bacterial Vaginosis Among a Cohort of Young Injection Drug Users in Baltimore, Maryland.

[19] P Feldblum, M Kuyoh, M Omari, K Ryan, J Bwayo, and M Welsh.Baseline STD prevalence in a community intervention trial of the female condom in Kenya.*Sex Transm Infect.* 2000 December; 76(6): 454456.

[20] Quigley MA, Weiss HA, Hayes RJ. Male circumcision as a measure to control HIV infection and other sexually transmitted diseases. *Curr Opin Infect Dis.* 2001 Feb; 14 (1): 715.

[21] Disease Risk - Top 10 Risk Factors for Acquiring an STD Oct 12, 2008 *http://std.about.com/od/riskfactorsforstds/tp/topriskfactors.htm*

[22] Wang QQ, Mabey D, Peeling RW, et al. Validation of syndromic algorithm for the management of genital ulcer diseases in China. *Int J STD AIDS*. 2002;13:469-74.

[23] Warner Lee et.al (2006):Condom Use and Risk of Gonorrhea and Chlamydia: A Systematic Review of Design and Measurement Factors Assessed in Epidemiologic Studies.

[24] Wi, T., V. Mesola, et al. (1998). "Syndromic approach to detection of gonococcal and chlamydial infections among female sex workers in two Philippine cities." *Sexually Transmitted Infections* 74(1S) supplement(1): 118S-122S.

[25] World Health Organization press release, WHO/64. *Sexually transmitted diseases three hundred and thirty-three million new, curable cases* in 1995. Geneva: World Health Organization, 1995