ADDIS ABABA UNIVERSITY
OFFICE OF THE GRADUATE PROGRAMMES

ASSOCIATION OF HIV INFECTION WITH SOME SELECTED FACTORS
and
MODELING THE CHANCE OF CONTRACTING HIV
(The Case of Awassa, SNNPR)

ALEMTSEHAI ABATE

JULY, 2006

ASSOCIATION OF HIV INFECTION WITH SOME SELECTED FACTORS and MODELING THE CHANCE OF CONTRACTING HIV

(The Case of Awassa, SNNPR)

BY

ALEMTSEHAI ABATE

DEPARTMENT OF STATISTICS

FACULTY OF SCIENCE

A THESIS SUBMITED TO THE OFFICE OF GRADUATE PROGRAMME, ADDIS ABABA UNIVERSITY, IN PARTIAL FULFILMENT FOR THE AWARD OF MASTER OF SCIENCE IN STATISTICS

ADDIS ABABA

JULY, 2006

ADDIS ABABA UNIVERSITY

OFFICE OF THE GRADUATE PROGRAMME

ASSOCIATION OF HIV INFECTION WITH SOME SELECTED FACTORS and MODELING THE CHANCE OF CONTRACTING HIV

(The Case of Awassa, SNNPR)

BY

ALEMTSEHAI ABATE

DEPARTMENT OF STATISTICS

FACULTY OF SCIENCE

APPROVRD BY:

ADVISOR:

PROFESSOR ESHETU WENCHEKO _____ 25 July 2006

BOARD OF EXAMINERS:

NAME	SIGNATURE	DATE
1. Dr. Bute Gotu	_____	25/07/06
2. Dr. Girma Taye	_____	25/07/06

CONTENTS

List of Tables:

ACKNOWLEDGEMENT

I would like to express my thanks from deep in my heart to my advisor, Professor Eshetu Wencheko, for his close and consistent follow up of this research undertaking and provision of constructive ideas and advise. Next, I am glad to thank all my instructors (of both undergraduate and postgraduate studies), without whose contribution, I might not be expected to be a statistician at all and carryout this study. Finally, many thanks should go to my family, relatives and friends who helped me in providing material and idea inputs to a successful completion of this study.

ABSTRACT

HIV/AIDS has become the serious health, economic, social and in general development problem worldwide. The epidemic is unique in its coverage of people from all racial groups, languages, genders, economic and academic status and the like.

Most past studies and researches in Ethiopia regarding the epidemic focused on awareness of people towards HIV, knowledge about condom distribution, pattern of condom distribution, knowledge of sexual behavior on HIV/AIDS and Sexually Transmitted Diseases, sexual violence and some related areas. But, to institute meaningful preventive measures for the control of HIV/AIDS, in addition to what have been done so far, there is a need to address the diversity of epidemics and make the prevention activity **evidence informed** through accurate epidemiological and behavioral information.

This study was aimed to explore evidences about the existence of HIV infection with some selected explanatory variables. A sample of 1461 cases, who voluntarily tested for HIV was taken from several Voluntary Counseling and Testing centers in Awassa city.

According to the findings of the study, there is evidence showing the presence of significant associations between HIV infection and explanatory variables like age, gender, marital status, occupation and educational level of individuals. This study reveals that the epidemic is almost evenly distributed through all religious groups and all the residence areas (urban and rural). Moreover, groups of people who are at a higher risk of HIV infection are identified in this study.

List of Acronyms Used in this Study:

AIDS: Acquired Immune Deficiency Syndrome.

CDC: Center for Disease Control and Prevention, USA.

CSA: Central Statistical Association, Ethiopia.

FAO: Food and Agriculture Organization, United Nations.

FGAE: Family Guidance Association, Ethiopia.

HIV: Human Immunodeficiency Virus.

ICA: Intelligence Community Assessment.

MOH: Ministry of Health, Ethiopia.

NGO: Non Governmental Organizations.

NIC: National Intelligence Council, USA.

OSSA: Organization for Social Service and AIDS.

PRB: Population Reference Bureau, United Nations.

SNNPR: Southern Nations, Nationalities and People's Region.

STD: Sexually Transmitted Diseases.

UNAIDS: The Joint United Nations Programme on HIV/AIDS.

UNICEF: United Nations Children's Fund.

USAID: United States Agency for International Development.

VCT: Voluntary Counseling and Testing, (for HIV).

WHO: World Health Organization, United Nations.

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

HIV infection is a viral infection caused by the human immunodeficiency virus (HIV), a virus that gradually destroys the immune system and results in infections that are hard for the body to fight. AIDS (Acquired Immune Deficiency Syndrome) is the final and most serious stage of HIV disease. Acquired Immunedeficiency Syndrome (AIDS) has killed more than 25 million people since it was first recognized in 1981, making it one of the most destructive epidemics in recorded history. Despite recent, improved access to antiretroviral treatment and care in many regions of the world, the AIDS epidemic claimed 3.1 million [2.8–3.6 million] lives in 2005; more than half a million (570 000) were children.

The total number of people living with HIV reached its highest level: an estimated 40.3 million [36.7–45.3 million] people are now living with HIV; women accounted for 46% of all adults living with HIV worldwide, and for 57% in Sub- Saharan Africa. Close to 5 million people were newly infected with the virus in 2005. Young people (15-24 years old) account for half of all new HIV infections worldwide- more than 6,000 become infected with HIV every day (UNAIDS/WHO, 2005).The HIV/AIDS pandemic continues to spread around the world at an alarming rate, and the number of people with the disease will grow significantly by the end of the decade, as it becomes more geographically diffuse (ICA 2002 -04 D by NIC).

The latest statistics on the world epidemic of AIDS & HIV published by UNAIDS/WHO in November 2005 shows the following. Their report gives the latest AIDS and HIV statistics for the whole world and for regions.

Table 1.1: World estimates of the HIV & AIDS epidemics at the end of 2005

		Estimate*	Range*
Number of people living with HIV/AIDS IN 2005	Total	40.3	36.7-45.3
	Adults	38.0	34.5-42.6
	Children	2.3	2.1-2.8
	Women	17.5	16.2-19.3
People newly infected with HIV IN 2005	Total	4.9	4.3-6.6
	Adults	4.2	3.6-5.8
	Children	0.70	0.63-0.82
AIDS death in 2005	Total	3.1	2.8- 3.6
	Adults	2.6	2.3- 2.9
	Children	0.57	0.51-0.67

*-millions

Source: UNAIDS/WHO AIDS epidemic update, December 2005

Note: The ranges around the estimates in this table define the boundaries within which the actual numbers lie, based on the best available information.

Table 1.2: Regional statistics for HIV & AIDS end of 2005

Region	Adults & children living with HIV*	Adults &children newly infected*	Adult infection Rate (%)	Death of adults and children*
Sub-Saharan Africa	25.8	3.2	7.2	2.4
East Asia	0.87	0.14	0.1	0.041
South and south-East Asia	7.4	0.99	0.7	0.48
Oceania	0.074	0.0082	0.5	0.0036
Eastern Europe & central Asia	1.6	0.27	0.9	0.062
Western & central Europe	0.72	0.022	0.3	0.012
North Africa and Middle East	0.51	0.067	0.2	0.058
North America	1.2	0.043	0.7	0.018
Caribbean	0.3	0.03	1.6	0.024
Latin America	1.8	0.2	0.6	0.066
Global Total	40.3	4.9	1.1	3.1

*millions

SOURCE: UNAIDS/WHO AIDS epidemic update, December 2005

According to the above report, by UNAIDS/ WHO, Sub-Saharan Africa is one of the most affected region in the world. This region has just over 10% of the world's population (PRB, 2005), but is home to more than 60% of all people living with HIV—25.8 million. In 2005, an estimated 3.2 million people in the region became newly infected, while 2.4 million adults and children died of AIDS. Among young people aged 15–24 years, an estimated 4.6% [4.2–5.5%] of women and 1.7% [1.3–2.2%] of men were living with HIV in 2005.

Ethiopia is one of the countries in Sub-Saharan Africa region, and among the highly affected countries by HIV/AIDS epidemic in the world. The HIV epidemic has evolved in Ethiopia from two reported AIDS cases in 1986 to a cumulative total of 147 000 by mid-2003. It is currently estimated that 1.5 million people are living with HIV and AIDS; about 96 000 are children aged under 15 years. Estimated national adult HIV prevalence in 2003 was 4.4% with uneven geographical distribution: 12.6% urban prevalence and 2.6% prevalence in rural settings; gender distribution is estimated at 3.8% male and 5% female. There were an estimated 539 000 AIDS orphans (children having lost one or both parents) in 2003; a cumulative total of 90 000 adults and 25 000 children had died of AIDS by end of 2003 (UNAIDS, 2004).

Though, no updated data found at national level for 2005, the 2004 report on the global AIDS epidemic shows the following about Ethiopia:

Table 1.3: HIV/AIDS Estimates (Ethiopia), end of 2003

Country HIV/AIDS Estimates, end 2003	
Adults (15-49) HIV prevalence rate	4.4%(range: 0.9% - 7.3%)
Adults (15-49) living with HIV	1400000 (range: 890000-2100000)
Adults and children(0-49) living with HIV	1500000(range:950000-2300000)
Women(15-49) living with HIV	770000(range:500000-12000000)
AIDS deaths (adults and children) in 2003	120000(range:74000-190000)

Source: 2004 Report on the global AIDS epidemic

Reliable statistics on HIV/AIDS are difficult or impossible to get for many countries. UNAIDS maintains the most comprehensive databases of information in the world on AIDS, but the UN organization relies on official government statistics of countries which experts believe sometimes understate the number of infected people (NIC, 2002)

The National Intelligence Council (NIC) of USA has launched a report titled "The Next Wave of HIV/AIDS" on the rising HIV/AIDS problem through 2010 in five countries: Nigeria, Ethiopia, Russia, India, and China. According to this report, adult prevalence rate in Ethiopia is estimated at between 10 and 18 percent which is the highest among the five countries, indicating that the disease has moved significantly into the general population. Adult prevalence is much higher in cities (13 to 20 percent) than in rural areas (5 percent). Heterosexual transmission is the primary mode of spread, and people with multiple partners especially those with sexually transmitted diseases (STDs) and prostitutes have significantly higher infection rates, ranging from 30 to 40 percent in STD-positive individuals to 50 to 70 percent in prostitutes.

1.2 Statement of the problem and Rationale of The study

Studies show that HIV prevention efforts work best when they are intensive, that is, comprehensive and long term. For example, intensive prevention programmes in the Mbeya region of Tanzania led to an increase in the use of condoms and the treatment of sexually transmitted infections between 1994 and 2000. Those changes were accompanied by a decline in HIV prevalence among 15–24 year-old women from 21% to 15% in the same period (Jordan-Harder et al., 2004).

According to a book titled "AIDS in AFRICA: Three Scenarios to 2025," the prevalence of HIV is different for men and women at different ages, and different for rural and urban populations. HIV prevalence probably also varies between rich and poor, educated and uneducated, employed and unemployed, but there are few statistics available so far that offer such breakdowns. The same source states that particularly for Africa, what happens over the next 20 years and beyond will depend on actions and decisions taken today, both on the continent itself and the rest of the world.

There is no single AIDS epidemic. Even within a country itself, epidemics can be extremely diverse (UNAIDS/WHO, 2005). Therefore, prevention strategies need to address the diversity of epidemics and must be evidence informed, through accurate epidemiological and behavioral information.

However, fundamental to all settings are comprehensive prevention strategies that include scale, intensity, consistency and sustainability as core requirements. All strategies must also recognize that HIV prevention and treatment are interlinked and that both should be simultaneously accelerated.

There are other basic approaches that can be applied to all HIV prevention efforts. First is the need to acknowledge that HIV prevention is a classic "public good" intervention that requires national governments to take the lead (including resource allocation) in building a strong response to the epidemic(UNAIDS/WHO, 2005).

Second is the need to ensure that all HIV prevention strategies take into account the growing linkages between AIDS and factors that put people at greater risk of HIV infection, such as age distribution, gender distribution, economical status (mostly reflected in one's occupation), educational status, marital status, etc...of specific populations. Testing the existence of statistically significant associations between the above factors and HIV infectivity can provide evidence for such evidence informed prevention mechanisms. Furthermore, modeling the chance (probability) of contracting HIV is helpful to identify a group with higher chance of getting the virus and take an evidence based prevention measure.

1.3 THE ROLE OF STATISTICS IN HIV/AIDS RESEARCH

Since the beginning of AIDS epidemic in developed countries statistical methods have played a considerable role in a variety of areas of research on HIV. These include the analysis of epidemiological patterns and studies of the natural history and clinical course of HIV-related diseases, the prediction of future trends and the design of trials and intervention and surveillance (Brookmeyer and Gail, 1986; May and Anderson, 1987). Statistical methods can also be used to make projections of AIDS and HIV case to few years in future, to estimate incubation period, to calculate number of people infected with HIV in past time intervals(Day et al., 1986; Cox et al., 1989).

Statistical models of the transmission dynamics of the virus have been of use in helping to interpret observed pattern and in dissecting the influence of different processes particularly components of sexual behaviors, on temporal changes in HIV prevalence and incidence.

Statistical methods have also played a role in the design and evaluation of interventions to limit the rate of spread of HIV and the analysis and interpretation of markers of the rate at which infected patient progress to AIDS (O'Brien et al., 1996) and many more ....

1.4 LITERATURE REVIEW

1.4.1 HIV Infection and Gender

In the HIV/AIDS epidemic, gender is defined as the array of societal beliefs, norms customs and practices that define "masculine" and "feminine" attributes and behaviors that play an integral role in determining an individual's vulnerability to infection, his or her ability to access care, support or treatment, and the ability to cope when infected or affected (WHO, 2006)

HIV infection is the most devastating new disease to have emerged in recent history. Although, worldwide, approximately as many women as men suffer from HIV, this aggregate figure conceals marked differences in the implications of the disease for men and women. Some of these result from biological differences in sex between men and women, but more result from socially defined gender differences (WHO, 2006).Such differences as:

- Women are probably more susceptible than men to infection from HIV in any given heterosexual encounter, due to biological factors – the greater area of mucous membrane exposed during sex in women than in men; the greater quantity of fluids transferred from

men to women; the higher viral content of male sexual fluids; and the microtears that can occur in vaginal (or rectal) tissue from sexual penetration. Young women may be especially susceptible to infection.

- Gender norms may also have an impact on HIV transmission. For example, in many places, gender norms allow men to have more sexual partners than women, and encourage older men to have sexual relations with much younger women. In combination with the biological factors cited above, this means that, in most places where heterosexual sex is the main mode of HIV transmission, infection rates are much higher among young women than among young men.
- Forced sex, which all too many women (and some men) experience at some point in their lives, can make HIV transmission even more likely, since it may result in more trauma and tissue tearing.

Around the world, from sub-Saharan Africa and Asia to Europe, Latin America and the Pacific an increasing number of women are being infected with HIV. It is often women with little or no income who are most at risk. In several southern African countries, more than three quarters of all young people living with HIV are women (WHO Regional Office for Africa, 2003; Reproductive Health Research Unit and Medical Research Unit, 2004), while in sub-Saharan Africa overall, young women between 15 and 24 years old are at least three times more likely to be HIV-positive than young men (UNAIDS, 2004).

According to "operational guide on gender and HIV/AIDS" (by UNAIDS 2005), in 1997, four out of ten people living with HIV/AIDS worldwide were women. The same source documented that by 2004, women made up almost 50% of people living with HIV/AIDS. In countries where

heterosexual transmission is the main mode of HIV transmission, women are more likely than men to be infected with HIV. In Sub-Saharan Africa, close to six out of ten adults (15-49 years) infected with HIV are women, and 75% of young people infected are women and girls. Throughout sub-Saharan Africa, HIV infection rates among teenage women are over five times higher than rates for teenage males. In Kenya, nearly one teenage woman in four is living with HIV, compared to one teenage male in 25 (UNAIDS, 1999).

According to UNAIDS, among young people aged 15–24 years, an estimated 4.6% [4.2–5.5%] of women and 1.7% [1.3–2.2%] of men were living with HIV in 2005.The highest "gender gap" in HIV infection rates is recorded between young women and men between 15-24 years old.

All the above figures show that female are being highly infected with HIV. The existence of significant association between gender and HIV infectivity is one question to be answered in this study.

1.4.2 HIV infection and Age

HIV/AIDS seriously affects adolescents throughout the world. One-third of all currently infected individuals are youth, ages 15 to 24, and half of all new infections occur in youth the same age. More than five young people acquire HIV infection every minute; over 7,000, each day; and more than 2.6 million each year (UNAIDS, 1999).

About 1.7 million new adolescent HIV infections over half of the world's total occur in sub-Saharan Africa. In fact, nearly 70 percent of people living with HIV/AIDS live in sub-Saharan Africa, and over 80 percent of AIDS deaths have occurred there(Akukwe 1999; Caldwell, 1997). Although HIV/AIDS rates vary considerably throughout sub-Saharan Africa generally lower in

western Africa and higher in southern Africa the epidemic has had a devastating effect on most African youth who often lack access to sexual health information and services. In particular, unmarried youth have great difficulty getting needed sexual health services. At the same time, cultural, social, and economic norms and pressures often put young African women at excess risk for HIV infection. In African countries with long, severe epidemics, half of all infected people acquire HIV before their 25th birthday and die by the time they turn 35 (UNAIDS 2000). The same report shows that in seven of 11 studies in Africa, at least one woman in five, ages 20 to 25, was HIV infected; most HIV-infected young women will not live to age 30.

In some African countries up to 60% of all HIV cases occur among 15-24 years old people (The World Bank, 2004).

Though it is rare, an emerging trend of rising infection rates among older generations in some countries may point to an important gap in prevention efforts with this age group (UNAIDS, 2005). According to this report, in South Africa, the rise in HIV prevalence among women older than 34 years is particularly striking and in Botswana, similar patterns are emerging among pregnant women aged 15–24, HIV infections have remained steady since 1999, but among their counterparts 25 years and older, prevalence has been rising constantly since 1992 and reached 43% when last measured in 2003. Infection levels among older men and women in Botswana were unexpectedly high: 29% for those 45–49 years old and 21% for those in their early 50s.

In the USA 10-15% of all reported new HIV infections occur among people over the age of 50, with a quarter of these among the over 60's. This amounted to around 78,000 people in April 2005(AIDS Info Net #616 April 2005), and the percentage of new infections occurring in this age group are rising. This is an increase of 18,000 people or 30% (AIDS Info Net #616 April

2004). In the UK, current data suggest that 8% of adults living with HIV or AIDS fall into the over 50 age category. Analysis of infection data collected from voluntary HIV testing and counseling centers in Uganda between 1999 and 2002 found that 4.6% of those who presented at the centers were older people. Of these 20% tested HIV positive (23.9% of women and 18% of men) (Clark, 2004).

In Ethiopia, studies show that HIV prevalence decreases with age (MOH, 2OO2; 2004). Youth aged 15 to 24 have the highest HIV prevalence. The same source indicated that the peak age for AIDS cases are 25 to 29 for both males and females. Given the average incubation period, between time of infection and the time of emergence of full blown disease is eight years, the mean age at which people become infected is probably 15 to 24 years.

1.4.3 The spread of HIV and Residence areas (Urban, Rural)

In developing countries estimates of HIV prevalence come chiefly from urban ser-sample. Data for rural populations are rare, and they are usually unrepresentative of the rural sector (Dyson, 2003). The levels of HIV infection are generally "significantly" high in urban areas .For example, sub-Saharan Africa (Caldwell and Anarfi, 1997) state that the urban level of HIV infection are typically four to ten times those of rural areas; (Carael, 1997) reports that rural HIV and STD prevalences have generally been found to be much lower than urban prevalences; and market towns tend to have a substantially higher occurrence of HIV than rural areas.

Though, the above findings state that the prevalence of HIV/AIDS is high in urban areas, the infection rate is still growing in rural areas too. According to the report by (UNAIDS, 2002), 7 million agricultural workers in 25 African countries have died of AIDS since 1985, and 16

million more deaths are likely in the next two decades. In 2001 alone, AIDS killed nearly 500,000 people in the six predominantly agricultural countries threatened with famine, most of whom were in their productive prime.

Although Ethiopia's national HIV prevalence rate is low (an estimated 4.4%) compared with many other countries to its south (Federal Ministry of Health Ethiopia, 2004), it faces many challenges in dealing with AIDS. The country's epidemic is concentrated mainly in urban areas, where HIV prevalence among pregnant women has averaged at 12–13% since the mid-1990s. In a society where some 85% of the population lives in rural areas, rising adult prevalence in rural areas (up from 1.9% in 2000 to 2.6% in 2003) gives cause for concern. Indeed, a large part of the AIDS burden is shifting to rural communities where more people are now being infected with HIV than in urban areas (Federal Ministry of Health Ethiopia, 2004).

1.4.4 HIV/AIDS and Educational Status

Educational levels make a huge difference of knowledge about transmission ways of HIV (UNICEF, 2004). For example, young women in Rwanda with secondary or higher education were five times as likely to know the main HIV transmission routes than were young women who with no formal education (Ministry of Health of Rwanda, 2001)

Most female sex workers originate from remote rural areas, are poorly educated and have little knowledge about HIV. Behavioral studies have shown that many sex workers continue to have unprotected sex even after discovering symptoms of sexually transmitted infections in themselves or their clients (Yang et al., 2005). Concerted efforts are needed to enable them to protect themselves against HIV and other sexually transmitted infections (Zhang et al., 2004).

One study of adolescents in 17 African countries showed that those with more education were far more likely to experience casual sex and to use condoms for casual sex when compared to less educated youth (UNAIDS REPORT, 2000).

The above reports and studies show that educational level has impact on awareness about HIV/AIDS. Does this mean, educated people are less infected with HIV? That is to mean, is their awareness accompanied with care to prevent HIV infection? This question needs to be answered in this study.

1.4.5 HIV/AIDS and Marital Status

The HIV/AIDS epidemic has been made possible by a number of factors. One of such factors is a higher level of sex outside marriage than old world agrarian societies (Caldwell, 2000). And another study shows that, in sub-Saharan Africa, the prevalence of HIV infection among young women is much higher than that among young men. Many women enter marriage HIV-infected, suggesting that men may be predominantly infected by their wives (Glynn, et al., 2003). This study outlines that at least one quarter of cases of HIV infection in recently married men were acquired from extramarital partnerships, and for both men and women, less than one half of cases of HIV infection were acquired from their spouse. In these sites, many infections in married men, even in those with HIV-infected wives, may be acquired from outside the marriage

In many countries, marriage, and women's own fidelity are not enough to protect them against HIV infection. Among women surveyed in Harare (Zimbabwe), in Durban and Soweto (South Africa), 66% reported having one lifetime partner, 79% had abstained from sex at least until the age of 17 (roughly the average age of first sexual encounter in most countries in the world). Yet,

40% of the young women were HIV-positive (Meehan et al., 2004). Many had been infected despite staying faithful to one partner. In Colombia, 72% of the women who tested HIV-positive at an antenatal site reported being in stable relationships. In India, a significant proportion of new infections is occurring in women who are married and who have been infected by husbands who (either currently or in the past) frequented sex workers. (UNAIDS, 2005)

Though it is unquestionable, that HIV prevalence is high among young and unmarried segment of any society, the above reports and studies indicate that, marriage is not guarantee for HIV free life. So, is there statistical association between marital status and HIV infection? This study tries to answer this question.

1.4.6 HIV/AIDS and Religion

Various reports and studies imply that HIV infection is visiting every religious group in the world. But, the extent of understanding and level of awareness differ at different religious societies. For instance, some faith traditions in Africa teach that AIDS is a shameful disease and a punishment for those who have been sexually promiscuous, and many adults are reluctant to admit to a disease that seems to imply promiscuity. One study showed three quarters of Nigerian Christian leaders believe that AIDS is a divine punishment (Caldwell, 2000)

Survey undertaken by Family Guidance Association of Ethiopia (FGAE, 1998) in Jimma area shows that religious affiliation of the youth appeared to play a role in the level of sexual activity. The proportion of sexually active females among Muslim was 9.7% followed by Protestants18.2% and orthodox Christians 23.9%. Whereas, the level of sexual activity among male Muslims was 45.5% which is lower by 1.7% than Orthodox Christians. Generally the findings show that those who are affiliated with Orthodox Christian are prone to premarital

sexual activity than the followers of other religions. Therefore it is important to establish whether HIV infection is religion dependent or not.

1.4.7 HIV/AIDS and Occupation

Occupation, in one or the other way, is the reflection of economic status of individuals and society at large.

Approximately 70 million young people are unemployed world wide. What is most important is the fact that, young people aged 15-24 have the highest infection rates from HIV/AIDS and other STDs (Women Care, 2001).

HIV/AIDS takes an especially heavy toll on the poor, because people may migrate in search of employment, or may look for rapid income, which can lead to high-risk behaviors such as drug abuse or involvement in prostitution. The consequences of poverty thus increase the risk of infection, and the disease in turn exacerbates poverty (FAO, 2001).

In the Ethiopian context, a study conducted in Awassa area by (Dejene Getahun, 2005) states that there is an association between economic level and sexual risk taking behavior. Can this statement lead us to conclude independence of HIV infection with economic level (specifically with occupation?). This question has given answer in Chapter three of this document.

1.5 OBJECTIVES OF THE STUDY

Most of the literature reveal that HIV infection rates are different with respect to different levels (categories) of those factors. So, the major and specific objectives of this study are the following:

General Objectives:

- ❖ To test whether HIV infection is associated with the factors discussed above.
- ❖ To develop a model that shows the level of chance of contracting HIV under certain conditions of the explanatory variables (factores).

Specific Objectives:

- ✓ To test whether there is a statistical dependence or not between HIV infection and the selected factors.
- ✓ To test whether those factors are affecting the infection level jointly or independently. That is, to test for conditional association between pairs of factors given the HIV infection status.
- ✓ To develop a statistical model that can be used in predicting the probability of individual's being infected under given conditions of the selected factors.
- ✓ To investigate which factor (among the selected ones) has a more significant effect on the response variable.

1.6 Applications of the Expected Results and the Study Rationale

1. Once we know the extent of relations (dependence) of HIV infection with those explanatory variables, it may help the concerned ones to allocate proportional resource to prevent and control the epidemic through those variables.
2. If we know that some of the explanatory variables are affecting the level of infection jointly, it may help the policy makers and the concerned organizations to act in such a way to invest the prevention resources to control one of the jointly acting factors, which indirectly control the effect of the partner factor and consequently the joint effect will be controlled. This may help in minimizing a required cost and resource.
3. The result can be used to identify and guide people who are at higher risk of infection (people who have relatively higher probability of being infected)
4. Much more applications can be extracted when the result is seen by area experts.

CHAPTER TWO

DATA AND STUDY METHODOLOGY

Input data for this study are obtained from five VCT centers (Bethezatha, OSSA, Family guidance association, youth center and Awassa Health center) in Awassa city.

2.1 SOME CHARACTERISTICS OF THE STUDY AREA

2.1.1 Location

Awassa, the capital of the Southern Nations, Nationalities and People's Regional State is located 7.06 degrees North of the Equator and 38.48 degrees of east. The city is at 1685 meters above sea level and situated 275 km south of Addis Ababa.

2.1.2 Population size

According to the 1994 census, the population size of the city of Awassa was 69,169 of which 50.6% were males and the remaining 49.4% were females (CSA, 1994). Table 2.1 shows the then population size by different age groups:

Table 2.1: Population size by five years age groups (1994)

Age group	Male		Female		Total
	N	%	N	%	
0-4	3938	51.3	3729	48.7	7667
5-9	4616	49.8	4657	50.2	9273
10-14	4762	47.0	5352	53.0	10114
15-19	4623	44.8	5692	55.2	10315
20-24	3881	47.6	4257	52.4	8138
25-29	3603	49.6	3657	50.4	7260
30-34	2701	55.3	2177	44.7	4878
35-39	2393	56.5	1841	43.5	4234
40-44	1641	65.2	874	34.8	2515
45-49	1064	65.0	572	35.0	1636
50-54	646	59.1	446	40.9	1092
55 &above	1161	56.7	886	43.3	2047

Source: CSA (1994)

2.1.3 Religion

The 1994 census shows that, Orthodox Christians were the majority constituting 65% of the total population of the city of Awassa, whereas Protestants, Catholics and Muslims constitute 26.9%, 3.3% and 4.0%, respectively (CSA, 1994).

Table 2.2: Population by Religion (1994)

	Orthodox	Protestant	Catholic	Muslim	Others
Both sex	44960	18604	2296	2826	390
Male	22644	9358	1116	1650	201
Female	22316	9246	1180	1176	189

Source: CSA (1994)

2.1.4 Literacy Status

According to the 1994 Census, out of all population aged ten years and above, 82.5% were all literate while the rest 17.5% were illiterate (see table 2.3).

Table 2.3: Population aged 10 years and above by literacy status (1994)

	Population size	**Illiterate**	**All literate**
Both sex	51996	9060	42872
Male	26248	3115	23103
Female	25748	5945	19769

Source: CSA (1994)

2.1.5 HIV Prevalence in the City of Awassa

Although, data on prevalence rate of HIV for near recent years are not obtained, data for 1998-2003 show the following. These data were taken from Awassa health center.

Table 2.4: Prevalence for Awassa Health Center (1998-2003)

Year	HIV prevalence rate at Awassa health center survilance site
1998	14.4
1999-2000	11.5
2001	10.0
2002	11.1
2003	8.8

Source: adopted from MOH 2004

2.2. DATA COLLECTION METHODES

The original (primary) data from which sample for this study is obtained were collected at different VCT centers in Awassa city using different approaches from volunteers. The large amount of data were collected using:

- **Interview method**: Face-to-face interview by councilor after getting counseled well.
- **Questionnaire method**: after providing the counseling service, a questionnaire was given to be filled by the volunteer.

The way data are collected makes this study a **Retrospective study,** because, information about factors which might be associated with HIV infection is obtained retrospectively for each person.

2.3 SAMPLING DESIGN AND TYPE OF SAMPLE

2.3.1 Sampling Design

Before taking sample for this study, the researcher has tried to consider different facts about the study population (population who voluntarily counseled and tested for HIV in Awassa city). Based on the background information gathered, it was decided to undertake non-random sampling scheme. Few reasons to not taking random sample:

- Some of the VCT centers have data for about 6 and 7 years. But if we consider the situation regarding HIV/AIDS in those distant past years, the awareness of people was not so developed as in recent years. So, the group of people who were being tested for

HIV during those far past years are not good representatives of the society. People were being tested then only for either marriage case or to go abroad. These people probably belong to the higher economic class and to relatively higher educational status. Moreover, there is no sampling frame containing all tested individuals in all of the VCT centers there. So, if simple random method selects people of those years, the sample will not be a good representative. And for lack of sampling frame, it cannot be employed.

- If we take cluster random sampling by considering different VCT centers as clusters, the sample will also lack representativeness. Because, people going to different VCT centers have different economic, academic, etc...backgrounds. For instance, people going to "Bethezatha" VCT center, where there is payment for HIV test, may belong to higher economic class, whereas people going to Awassa health center, where treatment is for free, may belong to the lower economic class. Again people going to "Youth center" may all be in the adolescent age group, whereas people going to "Family Guidance Association" may mostly be females and more of mothers.
- Stratified sampling faces almost all or some of the problems under simple random sampling and cluster sampling.
- If cost constraints is not imposed heavily on this study, systematic sampling after mixing all units from all VCT centers and preparing a sampling frame may give a good sample. But due to the budget and time constraints, this sampling scheme could not be employed.

Therefore, it was decided to take a sample using non-random method. As for me, even if I may have some centers with simpler and convenient type of data recording, the "convenience" and "quota" sampling methods may provide none representative sample.

Finally, "Judgment" sampling scheme was chosen based on the above constraints and nature of the available data. So, to get a representative sample, the recent 8 months (September, 2005-April, 2006) are chosen. But again for time and budget constraints, it was not possible to take sample of the eight months from all VCT centers. Instead, sample was taken for January, 2006 from Bethezatha, for September,2005 and December,2005 from Awassa health center, for November, 2005 - April, 2006 from Youth center, for February and March,2006 from OSSA, for October, 2005 and April, 2006 from Family guidance association. This selection was made based on data size for each month at those centers.

Note: Judgment sampling technique is the most appropriate if the population consists observations with **unequal importance**. This sampling scheme is used in this study in addition to budget constraint; observations from distant past years belong to certain group of society (not from all segment of society). Because, by then people were getting tested for HIV only either to go to abroad or to get married (this two norms involve only educated and youth groups in most common cases). But this study, which is going to use several explanatory variables, needs data which represent all groups of society. Actually, awareness towards HIV infection of the society has shown development recently than the past. Accordingly people getting tested for HIV recently are the better representative of the society than those in the past and VCTs also started to record several related documents to HIV infection recently. So, decision to use judgment sampling scheme is to get the most representative sample (the recent ones) and to omit inclusion of the most distant past observations, which do not represent the whole society.

Therefore, using the above method, the researcher has collected data on seven explanatory variables and one response variable from **1461** sample units. Sample sizes taken from each VCT center to aggregate 1461 are given in Table A8 in the Appendix.

Remark: Data are collected from several VCT centers due to the fact that people going to different VCT centers are with different backgrounds as mentioned above. And also the aim of this study is to investigate HIV infection rates at different groups of the society.

2.3.2 Sample type

Since the response variable, HIV test result, has only two possible outcomes (positive or negative), the sample collected is a **binary** sample. The independent variables, which are supposed to explain the response (dependent) variable, are more of categorical types having two or more levels each. Even those quantitative explanatory variables are also categorized to facilitate the test of association. As individuals are being counseled and tested independently, the binary responses of all cases are independent.

2.4 Methods for Data Analysis

This study aims to investigate the effects of those explanatory variables on the response variable, using two approaches, namely: Test of association and logistic regression analysis.

2.4.1 Test of Association

To test the existence of significant association between HIV infection and those selected factors, the *Pearson chi-square test statistic* will be employed. To test association between two variables

(factors), which came from same population and have different categories, the data should be presented in a $I \times J$ contingency table of observations n_{ij}, i = 1,2,....I, j = 1,2,...J

Some notations to be used in such contingency table of counts:

1. n_{ij} = the number of observations falling in the i^{th} row and j^{th} column

2. $n_{i.}$=marginal total of the i^{th} row, $n_{.j}$ = marginal total of the j^{th} column, and n.. = grand total

3. p_{ij} = the probability of having an observation fall in the i^{th} row and j^{th} column.

4. m_{ij} = the number of observations that one would expect to see in the i^{th} row and j^{th} column (based on some statistical model (hypothesis)).

Note: The marginal totals $p_{i.}$, $p_{.j}$, $m_{i.}$, $m_{.j}$ are defined like $n_{i.}$ and $n_{.j}$

Since the sample in this study is a **binary** sample with qualitative explanatory variables, the analysis of such sample begins by testing for independence of two factors. Two factors are independent if and only if $p_{ij} = p_{i.}\, p_{.j.}$ for all i and j.

So, to test for independence, we wish to test the model (hypothesis):

$$\mathbf{H_0}: p_{ij} = p_{i\cdot}\, p_{\cdot j},\ i = 1, 2, \ldots, I, j = 1, 2, \ldots J \qquad (2.1)$$

against the alternative

$\mathbf{H_A}$: model (2.1) is not true

Unbiased estimates of the marginal probabilities are:

$$\hat{p}_{i\cdot} = n_{i\cdot}/ n..$$

and $$\hat{p}_{\cdot j} = n_{\cdot j}/ n.. \qquad (2.2)$$

m_{ij} is commonly estimated as $m_{ij} = n..p_{ij}$.

If model (2.1) is true, we can estimate m_{ij} with

$$\hat{m}_{ij}^{(0)} = n..\ \hat{p}_{i\cdot}\ \hat{p}_{\cdot j} \qquad (2.3)$$

$$= n..(n_{i\cdot}/ n..)(n_{\cdot j}/ n..)$$

$$= n_{i\cdot}n_{\cdot j}/n.. ,$$

where the (0) in $\hat{m}_{ij}^{(0)}$ indicates that the estimate is obtained assuming that (2.1) holds. The Pearson chi-square test statistic is given by:

$$X^2_{cal} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}} \qquad (2.4)$$

which, if (2.1) is true and sample size is large, is approximately distributed as a $\chi^2((I-1)(J-1))$

H_0 in (2.1) is rejected at α level of significance if

$$X^2_{cal} > \chi^2(\alpha,(I-1)(J-1)) \qquad (2.5)$$

Note: A likelihood Ratio test statistic given by:

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\hat{m}_{ij}\log\left(\frac{\hat{m}_{ij}}{\hat{m}_{ij}^{(0)}}\right) \qquad (2.6)$$

can also be used to test association with degree of freedom

(I-1)(J-1).

Limitations:

1. Both X^2 and G^2 suffer from lack of accuracy when the expected cell counts are small (<5). However, for small samples X^2 is better than G^2 (Larntz, 1978). If the minimum expected cell count is about 1, χ^2 test based on Pearson test statistic X^2 works well (Fienberg, 1979). But, under all cases G^2 is more convenient to use in analyzing higher – dimension tables (Christensen, 1997).
2. If data are either unbalanced or sparse, test based on the asymptotic χ^2 - distribution will fail to work (Gajjar et al., 1998). In such cases we prefer the exact p-value. But, for large sample size, it needs long computation time. Therefore, for unbalanced and sparse data, exact p-value using Monte Carlo method is more appropriate.

2.4.2 Logistic Regression Analysis

This section is devoted to develop regression model for the log odds of a two category response variable, using multiple predictor (explanatory) variables.

Why logistic regression (why not OLS regression)?

There are several reasons to use logistic regression rather than the OLS regression to analyze the data:

- Use of dichotomous (binary) dependent variable in OLS regression violates the assumption of normality as a normal distribution is impossible with only two values.
- When the values can only be 0 or 1, residuals (errors) will be low for the portions of the regression line near Y=0 and Y=1, but high in the middle. Hence the error term will violate the assumption of homoscedasticity (equal variance) when dichotomy is used as a dependent. So, even with large samples, standard errors and significance tests will be in error.
- The assumption of linear relationship between the dependents and independents, by OLS, usually fails for many real data. But logistic regression doesn't assume this.
- OLS assumes that the independents must be quantitative in nature. But most of the independent variables in this study are qualitative.

Remarks: Logistic regression is preferred over the Probit Model. This is due to:

- The data in this study satisfy all assumptions of logistic regression such as binary response variable, qualitative or quantitative predictors, appropriate coding, large sample, independent errors and the like.
- The major assumption of Probit model is that it assumes error term has a standard normal distribution. But as mentioned above the error term may not be evenly distributed with constant variance. If this assumption fails, the probit model lacks accuracy in fitting data. So, Logistic model which does not assume such assumption is better to analyze the data in this study.

For the above and other reasons, logistic regression analysis is preferred to other generalized linear models.

For a binary response variable, the logistic transformation of success probability, p_i of the i^{th} individual can be modeled as a linear combination of k explanatory variables, x_{1i} , x_{2i} , ...x_{ki} , so that:

$$\log it(p_i)=\log\left(\frac{p_i}{1-p_i}\right)=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots\beta_k x_{ki} \qquad (2.7)$$

The explanatory variables in this study are:

A_i – age of the i^{th} individual

G_i – gender of the i^{th} individual

M_i – marital status of the i^{th} individual

O_i – occupation of the i^{th} individual

Rl_i – religion of the i^{th} individual

E_i – Educational level of the i^{th} individual

Rs_i – residence area of the i^{th} individual

To facilitate the analysis, each of these variables is categorized and coded as follows:

Age ($\leq$ 13 years=0, 14-30 years=1, 31-49 years=2,

50 years & above =3)

Gender (male =0, female =1)

Marital status (married =1, single including babies =2,

divorced =3, widowed =4)

Occupation (student =1, housewife =2, unskilled laborers &

unemployed =3, Merchant =4, Government & NGO employees =5,

Others (police military, drivers,....) =6)

Religion (Orthodox =0, protestant =1, Muslim = 2, Catholic =3,

Others including babies =4)

Education (Illiterate=0, Primary (Kg-8) =1, secondary (9-12) =2,

higher (10+1 & above) =3)

Residence area (urban=0, rural =1)

Note: The term babies in this study refer to cases who are aged 13 years and below. There are 21 babies in this study which is only 1.4% of the sample. So, inclusion of this figure under those categories will not affect the result much.

Then, using these explanatory variables, the logistic regression of success probability (probability of contracting HIV, that is, having an HIV positive test result) of the form:

$$\log it(p_i) = \beta_0 + \beta_1 A_i + \beta_2 G_i + \beta_3 M_i + \beta_4 O_i + \beta_5 RI_i + \beta_6 E_i + \beta_7 Rs_i \qquad (2.8)$$

is to be fitted.

2.4.2.1 Test of Goodness of Fit

After fitting a model, it is natural to enquire about the extent to which the fitted values of the response variable under the model compare with observed values. The goodness of fit of the model in this study is to be tested in either of the following two approaches (The choice of test type is confounded on software package available).

a. Deviance analysis

This is based on the likelihood function of the observed $\hat{p}_i$ for the fitted model (current model), say $\hat{L}_c$, and the likelihood function for the true success probability under the assumed perfect model (full or saturated model), say $\hat{L}_f$. The deviance denoted by D is given by:

$$D = -2\log\left(\hat{L}_c/\hat{L}_f\right) = -2\left[\log\hat{L}_c - \log\hat{L}_f\right] \qquad (2.9)$$

Larger values of D are encountered when $\hat{L}_c$ is small relative to $\hat{L}_f$, indicating that the current model is poor one. So, to test goodness of fit, we can use the deviance (change in -2Log (likelihood)).

b. Pearson's X^2 –statistic

An alternative approach to test goodness of fit is to use Pearson's X^2 –statistic defined by:

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - n_i\hat{p}_i)}{n_i\hat{p}_i(1-\hat{p}_i)}, \qquad (2.10)$$

where y_i= number of successes in i^{th} category

n_i= number of individuals in i^{th} category

p_i = success probability in i^{th} category

Note: 1. Both the deviance and X^2 –statistic have the same asymptotic χ^2-distribution.

2. Since the maximum likelihood estimates of the success probabilities maximize the likelihood function for current model, the deviance is the goodness of fit statistic that minimized by these estimates. On this basis, it is more appropriate to use the deviance.

3. All the analyses using the above methods are to be done using an SPSS package.

Limitation: For ungrouped binary data, with $n_i = 1$, $i = 1, 2, \ldots, n$, the deviance depends on only the fitted success probability p_i and so is uninformative about the goodness of fit of **overall model** (Collett, 1991). However, it can still be used as the best basis for comparing models. That is, to test the importance of including or removing **particular independent variable**. The test statistic (2.10) faces similar problem. So, if time (consequence of budget problem) will not allow me to group the data under all the explanatory variables, I will opt to use other test statistics as **Hosmer and Lemeshow's** goodness of fit test (Hosmer and Lemeshow, 2000) , **Psuedo R^2** based on appropriateness for data to test the goodness of **overall model**. These alternative goodness of fit tests are defined as:

i. Hosmer-Lemeshow Test. In this approach, data are divided into 10 groups. From each of such group, the observed and expected number of events will be computed. Then the test statistic is given by:

$$\hat{C} = \sum_{k=1}^{10} \frac{(O_k - E_k)^2}{v_k} \qquad (2.11)$$

where, O_k=observed number of events in k^{th} group

E_k=expected number of events in k^{th} group

v_k=a variance correction factor for the k^{th} group

ii. **Pseudo-R^2.** Deviance can be thought of as a measure of how poorly the model fits (i.e., lack of fit between observed and predicted values), an analogy can be made to

the sum of squares residual in ordinary least squares. In SPSS, there are two modified versions of this basic idea, one developed by

Cox & Snell and the other developed by Nagelkerke. The Cox and Snell R-square is computed as follows:

Cox & Snell Pseudo-R^2 :

$$R^2 = 1 - \left[\frac{-2LL_{null}}{-2LL_k} \right]^{2/n} \qquad (2.12)$$

Because this R-squared value cannot reach 1.0, Nagelkerke modified it such that the correction increases the Cox and Snell version to make 1.0 a possible value for R-squared.

Nagelkerke Pseudo-R^2:

$$R^2 = \frac{1 - \left[\frac{-2LL_{null}}{-2LL_k} \right]^{2/n}}{1 - \left(-2LL_{null}\right)^{2/n}} \qquad (2.13)$$

where, the null model is the logistic model with just the constant and the *k* model contains all the predictors in the model.

-2LL stands for -2 times log likelihood.

CHAPTER THREE

RESULTS AND DISCUSSION

3.1 Test of Association

In this section we test whether the explanatory variables have statistically significant association with HIV infection or not.

General remark: Because the study is associated with health problem, it becomes essential to minimize the chance of occurrence of **Type I** error to the possible extent. So, all of the tests of association below are carried out at 1% level of significance.

3.1.1 HIV Infection and Age

The contingency table of the 1461 case with regard to age and HIV infection is given in Table 3.1.

Table 3.1: Age *Test result Cross tabulation

			Result		
			Negative	Positive	Total
Age	13 years and bellow	Count	20	1	21
		Expected Count	18.4	2.6	21.0
	14- 30 years	Count	1086	116	1202
		Expected Count	1055.6	146.4	1202.0
	31-49 years	Count	165	56	221
		Expected Count	194.1	26.9	221.0
	50 years and above	Count	12	5	17
		Expected Count	14.9	2.1	17.0
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0

The hypothesis to be tested is:

$\mathbf{H_0}$: HIV infection and age of an individual are independent.

$\mathbf{H_A}$: HIV infection does not depend on the age of an individual.

Though, the sample size is large, it was observed that the expected cell counts for two cells are less than 5. Under such cases, the assumption that X^2 and G^2 have an asymptotic χ^2-distribution will lack accuracy. So, we need an exact p-value. But, for this large sample size, the exact p-values can not be computed. Therefore, the best approach is to use the Monte Carlo estimate of p-value. This value generated from StatXact package is:

```
Statistic based on the observed 4 by 2 table(x):
     CH(X): Pearson Chi-Square Statistic =       48.76

    Pr {CH(X) .GE.        48.76} =      0.0000

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        48.76} =      0.0000
     99.00% Confidence Interval     = (      0.0000,      0.0005)
```

Both the point estimate and the 99% confidence interval estimates show that there is sufficient evidence to reject the null hypothesis almost at all levels of significance. This implies that HIV infection is associated with age group of individuals.

Having this conclusion we may be interested to identify which age group is highly affected by HIV. This can be seen from the percentage (within-age) of infected people. *Table A1,* in the Appendix, shows that 4.8% of age 13 years and below, 9.7% of age 14-30 years, 25.3% of age 31-49 years, 29.4% of age 50 years and above are infected with HIV.

Note (General Limitation): The above percentage figures may suffer from lack of precision. It should be interpreted carefully. Because these figures are calculated for cases who made voluntary test for HIV. But, due to the awareness gap, most of the people at older age groups may (usually) not be tested unless they get sick or the like. So, many of the old aged people tested may have got treated after getting sick and had an HIV positive test result. This may raise the percentage of HIV positives in old age groups.

3.1.2 HIV Infection and Gender

Data on gender and test result of sampled cases are given below.

Table 3.2: Gender *Test result Cross tabulation

			Result		Total
			negative	positive	
Gender	male	Count	679	65	744
		Expected Count	653.4	90.6	744.0
	female	Count	604	113	717
		Expected Count	629.6	87.4	717.0
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0

The hypothesis to be tested here is:

H_0: HIV infection has no association with gender.

H_A: HIV infection and gender are associated.

As can be seen from the Table 3.2, the data are balanced and no cell has expected count less than 5. So, both X^2 and G^2 have asymptotic χ^2-distribution.

The SPSS output with asymptotic p-values is given in Table 3.3.

Table 3.3: Chi-Square Tests (Gender vs Test result)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square(X^2)	16.835	1	.000
Likelihood Ratio(G^2)	16.993	1	.000

Both the X^2 and G^2 statistics lead to a rejection of H_0. So, HIV infection and gender have a statistical association. *Table A2*, of the Appendix shows that 15.8% of females and 8.7% of male tested voluntarily are HIV positive.

This result leads us to infer that females are the highly affected group as compared to males. Because, the data show that there is no much awareness gap between males and females as there is no much difference in number of males and females treated.

3.1.3 HIV infection and Marital status

Table 3.4: **Marital status Versus Result Cross tabulation**

			Result		Total
			negative	positive	
Mar.status	Married	Count	200	65	265
		Expected Count	232.7	32.3	265.0
	single(unmarried)	Count	990	39	1029
		Expected Count	903.6	125.4	1029.0
	divorced(separated)	Count	72	44	116
		Expected Count	101.9	14.1	116.0
	Widowed	Count	21	30	51
		Expected Count	44.8	6.2	51.0
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0

We test,

$\mathbf{H_0}$: HIV infection has no association with marital status against

$\mathbf{H_A}$: HIV infection is associated with marital status of individuals.

To test this hypothesis, Pearson's X^2 having asymptotic χ^2-distribution is appropriate, because no cell has expected count less than 5. The SPSS output is given bellow.

Table 3.5: Chi- Square Tests (Marital status vs Test result)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square(X^2)	281.068(a)	3	.000
Likelihood Ratio(G^2)	232.660	3	.000

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.21.

Using the usual reasoning, this result also shows that the data provides sufficient evidence to reject the null hypothesis, implying that HIV infection is significantly associated with marital status of people. From *Table A3*, in the Appendix, we see that 24.5% of married, 3.8% of singles, 37.9% of divorced and 58.8% of widows tested voluntarily are HIV positives.

Note: These percentage figures are also subjected to limitations mentioned in section 3.1.1 (awareness gap between singles and other categories). So, people who are married, divorced, and widowed may not (as commonly seen) be aware and get voluntarily tested for HIV unless they feel some symptoms.

3.1.4 HIV infection and Residence area

Data on HIV infection status classified by residence area of cases are given below.

Table 3.6: **Residence area* Test Result Cross tabulation**

			Result		
			negative	positive	Total
Residence	urban	Count	1120	156	1276
		Expected Count	1120.5	155.5	1276.0
	Rural	Count	163	22	185
		Expected Count	162.5	22.5	185.0
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0

We test: $\mathbf{H_0}$: HIV Infection is independent of Residence area against

$\mathbf{H_A}$: HIV Infection depends on Residence area.

The X^2 and G^2 –statistics having asymptotic χ^2-distribution (for the data are balanced), the SPSS output are:

Table 3.7: Chi-Square Tests (Residence area vs Test result)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.017	1	.897
Likelihood Ratio	.017	1	.896

Both the X^2 and G^2 show that there doesn't exist a significant association between HIV infection and residence area.

Hence we don't reject the null hypothesis.

Table A4, in the Appendix, shows that the infection rate is 12.2% in urban areas and 11.9% in rural areas. These figures show that infection rates are almost same at urban and rural areas. But, the within result percentages (in the same table) show that there 87.6% of infected people are from urban and the rest 12.4% are from rural areas. Therefore, the infection rates given above should be interpreted with the general limitation of Section 3.1.1. Because, there is high awareness gap at rural areas as compared to urban areas.

Note: People at rural areas have no VCT centers near to their residence area and also there is relatively lower awareness about HIV/AIDS. So, they are very few people (most probably the ones who feel some symptoms of HIV/AIDS) that get tested for HIV from there. Therefore the above conclusion and given proportions should consider this limitation.

3.1.5 Association of HIV infection with Occupation

Data on HIV test results classified according to occupation of individuals are given below.

Table 3.8: Occupation * Test Result Cross tabulation

			Result		
			negative	positive	Total
Occupation	Student	Count	478	13	491
		Expected Count	431.2	59.8	491.0
	Housewife	Count	82	49	131
		Expected Count	115.0	16.0	131.0
	unskilled lab&unemployed	Count	203	61	264
		Expected Count	231.8	32.2	264.0
	Merchant	Count	168	20	188
		Expected Count	165.1	22.9	188.0
	Gov&NGO employee	Count	194	13	207
		Expected Count	181.8	25.2	207.0
	Others (includes police, military, drivers...)	Count	158	22	180
		Expected Count	158.1	21.9	180.0
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0

We test; $\mathbf{H_0}$: HIV infection is independent of one's occupation against

$\mathbf{H_A}$: HIV infection depends on one's occupation

Since the data in Table 3.8 are balanced (no cell has expected count less than 5), the asymptotic χ^2-distribution will give us accurate p-values.

Table 3.9: Chi-Square tests (Occupation vs Test result)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	156.215(a)	5	.000
Likelihood Ratio	145.875	5	.000

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 15.96.

The p-values are so small (almost null). This shows that there is sufficient evidence to reject H_0. Consequently, we conclude that HIV infection is significantly associated with occupation. The

rate of infection (based on the sampled data) at different occupation categories are: 2.6% of students, 37.4% of housewives, 23.1% of unskilled laborers, 10.6% of merchants, 6.3% of governmental and non governmental organization employees, 12.2% of others (including police, military, drivers...) are infected.

3.1.6 Association of HIV infection with Religion

Table 3.10: Religion * Test Result Cross tabulation

			Result		
			negative	positive	Total
Religion	Orthodox Christian	Count	666	107	773
		Expected Count	678.8	94.2	773.0
	Protestant	Count	440	59	499
		Expected Count	438.2	60.8	499.0
	Muslim	Count	134	7	141
		Expected Count	123.8	17.2	141.0
	Catholic	Count	20	2	22
		Expected Count	19.3	2.7	22.0
	Others(including babies)	Count	23	3	26
		Expected Count	22.8	3.2	26.0
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0

$\mathbf{H_0}$: HIV infection is independent of one's religion against

$\mathbf{H_A}$: HIV infection depends on one's religion.

The data table above shows that there are some cells with expected cell count less than 5. So, asymptotic p-values may lead to a wrong conclusion. Therefore, the better way is to use the exact p-value using Monte Carlo's approach.

```
StatXact-4 Output
CHI-SQUARE TEST FOR INDEPENDENCE
Statistic based on the observed 5 by 2 table(x):
```

```
     CH(X): Pearson Chi-Square Statistic =       9.123

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        9.123} =      0.0621
     99.00% Confidence Interval     = (      0.0559,      0.0683)
```

Both the point and interval estimates of the exact p-value show that H_0 should not be rejected at 5% or less significance levels. Consequently, we conclude that HIV infection doesn't depend on one's religion. This can also be seen from *Table A6*, in the Appendix. About 13.8% of Orthodox, 11.8% of Protestants, 5% of Muslims, 9.1% of Catholic, and 11.5% of others are infected.

3.1.7 Association of HIV infection with Educational level

Table 3.11: Education * Test Result Cross tabulation

			Result		
			negative	positive	Total
Education	Illiterate	Count	111	42	153
		Expected Count	134.4	18.6	153.0
	primary(KG-8)	Count	324	79	403
		Expected Count	353.9	49.1	403.0
	Secondary(9-12)	Count	486	47	533
		Expected Count	468.1	64.9	533.0
	Higher(10+1 &above)	Count	362	10	372
		Expected Count	326.7	45.3	372.0
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0

H_0: There is no association between HIV infection and one's educational level against

H_A: HIV infection has association with one's educational status.

To test this hypothesis, we can safely use the asymptotic χ^2-distribution, because all cell contents are larger than 5.

Table 3.12: Chi-Square Tests (Education vs Test result)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	91.060(a)	3	.000
Likelihood Ratio	94.055	3	.000

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 18.64.

These results (output of SPSS) show that the null hypothesis should be rejected at almost all levels of significance. Thus; HIV infection depends significantly on one's educational status.

Referring to *Table A7*, in the Appendix, we see that 27.5% of illiterates, 19.6% of those who attend up to primary school, 8.8% of those who attended secondary school, and 2.7% of those who attended higher institutes are infected with HIV.

3.2 Conditional Test of Association among explanatory variables

This section is devoted to test whether each pair of explanatory variables (factors) has a compound (joint) effect on HIV infection or each of the variables has an independent effect, given that a person is HIV positive. To conduct test of this section, a sample of 178 HIV positive cases (total HIV positives in the data) were taken. So, these tests are conditional test of association given the HIV infection status.

Testing for independence using Monte Carlo approach (since almost all data tables are sparse and/or unbalanced), of the explanatory variables for those HIV positive cases, reveal the following results.

The test for age versus marital status, age versus religion, age versus educational level, gender versus residence area, gender versus religion, gender versus educational level, marital status versus residence area, marital status versus religion, residence area versus occupation, residence

area versus religion, residence area versus educational level, religion versus occupation, and religion versus educational level show that these pairs of variables have no associated effect on HIV infection at 1% level of significance. While test of association on age versus gender, age versus residence area, age versus occupation, gender versus marital status, gender versus occupation, gender versus residence area, marital status versus occupation, marital status versus education, and educational level versus occupation reveals that these pairs of explanatory variables have associated (joint) effect on HIV infection at 1% level of significance.

The Monte Carlo approach outputs of StatXact for the above tests are given below:

Age vs Gender

```
StatXact-4 Output

CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       30.17

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        30.17} =      0.0000
     99.00% Confidence Interval     = (      0.0000,      0.0005)
```

Age vs marital status

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 4 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       13.98

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        13.98} =      0.1072
     99.00% Confidence Interval     = (      0.0992,      0.1152)
```

Age vs occupation

```
StatXact-4 Output

CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 6 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       37.14

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        37.14} =      0.0016
     99.00% Confidence Interval     = (      0.0006,      0.0026)
```

Age vs Residence area

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 6 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       37.14

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        37.14} =      0.0022
     99.00% Confidence Interval     = (      0.0010,      0.0034)
```

Age vs Religion

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 5 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       68.29

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        68.29} =      0.0149
     99.00% Confidence Interval     = (      0.0118,      0.0180)
```

Age vs Educational level

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 4 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       11.21

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        11.21} =      0.2060
     99.00% Confidence Interval     = (      0.1956,      0.2164)
```

Gender vs marital status

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       13.64

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        13.64} =      0.0036
     99.00% Confidence Interval     = (      0.0021,      0.0051)
```

Gender vs residence area

Table 3.13: Chi-Square Tests (Gender vs Residence area)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13.637	3	.003
Likelihood Ratio	14.609	3	.002

0 cells (.0%) have expected count less than 5. The minimum expected count is 10.96.

Note: Since data in this case are balanced type, an asymptotic p-value is better than the exact one. So, the SPSS output given above is using the asymptotic approach.

Gender vs occupation

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 6 table(x):
     CH(X): Pearson Chi-Square Statistic =       47.66

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        47.66} =      0.0000
     99.00% Confidence Interval     = (      0.0000,      0.0005)
```

Gender vs religion

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 5 table(x):
     CH(X): Pearson Chi-Square Statistic =       2.104

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        2.104} =      0.7605
     99.00% Confidence Interval     = (      0.7495,      0.7715)
```

Gender vs educational level

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       7.856

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        7.856} =      0.0496
     99.00% Confidence Interval     = (      0.0440,      0.0552)
```

Mart status vs Residence area

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       3.764

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        3.764} =      0.2969
     99.00% Confidence Interval     = (      0.2851,      0.3087)
```

Mart status vs occupation

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 6 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       57.48

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        57.48} =      0.0000
     99.00% Confidence Interval     = (      0.0000,      0.0005)
```

Marital status vs religion

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 5 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       13.57

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        13.57} =      0.3214
     99.00% Confidence Interval     = (      0.3094,      0.3334)
```

Marital status vs educational level

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 4 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       28.68

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        28.68} =      0.0006
     99.00% Confidence Interval     = (      0.0000,      0.0012)
```

Occupation vs residence area

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 6 table(x):
     CH(X) : Pearson Chi-Square Statistic =       7.268

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        7.268} =      0.1901
     99.00% Confidence Interval     = (      0.1800,      0.2002)
```

Residence area vs religion

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 5 table(x):
     CH(X): Pearson Chi-Square Statistic =       14.11

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        14.11} =      0.0280
     99.00% Confidence Interval     = (      0.0238,      0.0322)
```

Residence area vs educational

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 2 by 4 table(x):
     CH(X): Pearson Chi-Square Statistic =       2.429

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        2.429} =      0.4874
     99.00% Confidence Interval     = (      0.4745,      0.5003)
```

Occupation vs religion

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 5 by 6 table(x):
     CH(X) : Pearson Chi-Square Statistic =       26.12

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        26.12} =      0.1616
     99.00% Confidence Interval     = (      0.1521,      0.1711)
```

Occupation vs educational level

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 4 by 6 table(x):
     CH(X): Pearson Chi-Square Statistic =       47.58

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        47.58} =      0.0002
     99.00% Confidence Interval     = (      0.0000,      0.0006)
```

Religion vs educational level

```
CHI-SQUARE TEST FOR INDEPENDENCE

Statistic based on the observed 4 by 5 table(x):
     CH(X): Pearson Chi-Square Statistic =       15.73

Monte Carlo estimate of p-value:
     Pr {CH(X) .GE.        15.73 } =      0.1964
     99.00% Confidence Interval     = (      0.1862,      0.2066)
```

3.3 Logistic Regression Analysis

To study the effects of those seven explanatory variables on a binary response variable, namely HIV test result, having two outcomes only (positive =1, or negative= 0), the logistic regression will be used. In this study, success outcome is the HIV positive test result. Then log odd of contracting HIV is to be expressed as linear combination of those selected predictor variables. To compute odds of contracting HIV for any category, we always need to have a reference category.

SPSS package takes automatically either the last (in order of coding) or the first category as reference category. Then the odd of success for any category will be interpreted in relation to the reference category selected.

The SPSS codlings for the explanatory variables in this study are given in Table 3.14.

Table 3.14: Categorical Variables Codings

			Parameter coding				
		Frequency	(1)	(2)	(3)	(4)	(5)
Occupation	Student	491	1.000	.000	.000	.000	.000
	Housewife	131	.000	1.000	.000	.000	.000
	uskilled lab&unemployed	264	.000	.000	1.000	.000	.000
	Merchant	188	.000	.000	.000	1.000	.000
	Gov&NGO employee	207	.000	.000	.000	.000	1.000
	Others(inclusdes police, military, drivers...)	180	.000	.000	.000	.000	.000
Religion	Orthodox Christian	773	1.000	.000	.000	.000	
	Protestant	499	.000	1.000	.000	.000	
	Muslim	141	.000	.000	1.000	.000	
	Catholic	22	.000	.000	.000	1.000	
	Others(including babies)	26	.000	.000	.000	.000	
Education	Illiterate	153	1.000	.000	.000		
	primary(KG-8)	403	.000	1.000	.000		
	Secondary(9-12)	533	.000	.000	1.000		
	Higher(10+1 &above)	372	.000	.000	.000		
Mar.status	Married	265	1.000	.000	.000		
	single(unmarried)	1029	.000	1.000	.000		
	divorced(separated)	116	.000	.000	1.000		
	Widowed	51	.000	.000	.000		
Age	13 years and below	21	1.000	.000	.000		
	14- 30 years	1202	.000	1.000	.000		
	31-49 years	221	.000	.000	1.000		
	50 years and above	17	.000	.000	.000		
Gender	Male	744	1.000				
	Female	717	.000				
Residece	Urban	1276	1.000				
	Rural	185	.000				

Making use of the new coding, logistic regression coefficients can be estimated using the maximum likelihood estimation method. This will be done using the SPSS package. All the estimated coefficients for a model may not be significant showing that some of the explanatory variables may be insignificant and irrelevant to explain the response variable. Therefore, we need to select the most appropriate model.

3.3.1 Model Selection

In examining the effect of including terms in, or excluding terms from, a model, we consider the change in Deviance. The change in the deviances of two nested models measures the extent to which the additional term(s) improve the fit of the model to the observed response variable. To select appropriate model for this study, it may be needed to observe deviances of $2^7=128$ different models containing only the main effects (for 7 explanatory variables) manually, which becomes a cumbersome task. There are statistical packages which give automatic results regarding the variables to be included in the model. But these have also their own limitations:

Limitation: Forward selection and backward elimination procedures, designed in statistical packages, have a very arbitrary nature. So, their use should be avoided in model selection (Collett, 1991).

Therefore, in this study, variable selection is based on the significance of association they have with response variable (based on tests in Section 3.1) and the corresponding deviance analysis. This approach may lead to a better model selection, because it is free of both subjective bias (if selection were subjective) and of those problems due to arbitrariness of software programmes (if forward or backward elimination methods were used) for logistic regression analysis.

We say two models are nested if one model contains additional terms which belong to the other model.

As can be seen from Section 3.1, association of HIV infection is highly associated with age, gender, marital status, occupation, and educational level. So, let's start fitting logistic model by stepwise inclusion of these and other explanatory variables as well. The SPSS outputs followed by statistical tests are given below:

Step1: Model with the first explanatory variable (Age):

Table 3.15: Iteration History(a,b,c) (for estimating constant term)

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	1120.323	-1.513
	2	1083.473	-1.909
	3	1082.788	-1.974
	4	1082.788	-1.975
	5	1082.788	-1.975

a Constant is included in the model.
b Initial -2 Log Likelihood: 1082.788
c Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Table 3.16: Variables in the Equation (model with only constant term)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.975	.080	609.827	1	.000	.139

At this block (sub step) the constant for the model is estimated, and is β_0 = -1.975. The deviance of having a model with only the constant is 1082.788, which shows that this model is too far from a full model.

Block 1: Method = Enter

Table 3.17: Variables in the Equation (model with age as independent variable)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Age			44.784	3	.000	
	Age(1)	-2.120	1.155	3.372	1	.066	.120
	Age(2)	-1.361	.541	6.326	1	.012	.256
	Age(3)	-.205	.554	.137	1	.711	.815
	Constant	-.875	.532	2.705	1	.100	.417

a Variable(s) entered on step 1: Age.

At this block, the coefficients of each category of age are estimated. The deviance of this model including age is 1041.702, indicating the model is still too far from the full model. To test the importance of including age to the model, in addition to that constant, the difference in deviances is used. That is, 1082.788 - 1041.702 = 41.086, which is significant showing inclusion of age in the model is important (see table 3.19 below).

Table3.18: Model Summary (for model at step 1)

Step	-2 Log likelihood
1	1041.702(a)

a Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table 1.19: Omnibus Tests of Model Coefficients (for model at step 1)

		Chi-square	df	Sig.
Step 1	Step	41.086	3	.000
	Block	41.086	3	.000
	Model	41.086	3	.000

Using similar reasoning and SPSS outputs for the rest of explanatory variables, we proceed as follows:

Step 2: Inclusion of Gender, in the presence of age to the model

Table 3.20: Variables in the Equation (for model at step 2)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Age			55.924	3	.000	
	Age(1)	-2.270	1.164	3.801	1	.051	.103
	Age(2)	-1.500	.554	7.344	1	.007	.223
	Age(3)	-.131	.566	.054	1	.817	.877
	Gender(1)	-.898	.175	26.401	1	.000	.407
	Constant	-.386	.551	.492	1	.483	.680

a Variable(s) entered on step 2: Gender.

Table 3.21: Model Summary (for model at step 2)

Step	-2 Log likelihood
1	1013.769(a)

a Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table 3.22: Omnibus Tests of Model Coefficients (for model at step 2)

		Chi-square	df	Sig.
Step 1	Step	27.934	1	.000
	Block	27.934	1	.000
	Model	69.019	4	.000

The deviance due to inclusion of gender in the presence of age is 1013.769. So, the difference in deviance is 1041.702 -1023.769 = 27.934 (see Table 3.22). This also shows including gender in the model at Step 1 has significant importance to explain the response variable.

Step 3: Inclusion of Occupation to model at Step 2

Table3.23: Variables in the Equation (for model at step 3)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Age			24.890	3	.000	
	Age(1)	-1.471	1.195	1.514	1	.219	.230
	Age(2)	-.589	.573	1.057	1	.304	.555
	Age(3)	.406	.587	.478	1	.489	1.501
	Gender(1)	-.417	.207	4.070	1	.044	.659
	Occupation			82.017	5	.000	
	Occupation(1)	-1.518	.381	15.910	1	.000	.219
	Occupation(2)	1.128	.337	11.192	1	.001	3.090
	Occupation(3)	.712	.290	6.039	1	.014	2.038
	Occupation(4)	-.232	.337	.473	1	.492	.793
	Occupation(5)	-.867	.374	5.377	1	.020	.420
	Constant	-1.339	.642	4.344	1	.037	.262

a Variable(s) entered on step 3: Occupation.

Table 3.24: Model Summary (for model at step 3)

Step	-2 Log likelihood
1	910.977(a)

a Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table 3.25: Omnibus Tests of Model Coefficients (for model at step 3)

		Chi-square	df	Sig.
Step 1	Step	102.791	5	.000
	Block	102.791	5	.000
	Model	171.811	9	.000

This model results in a very significant reduction in deviance, 102.791, showing inclusion of occupation to the model at Step 2 has great importance in explaining the response variable.

Step 4: Inclusion of Education to the Model at Step 3.

Table 3.26: Variables in the Equation (for model at step 4)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Age			24.484	3	.000	
	Age(1)	-1.622	1.193	1.847	1	.174	.198
	Age(2)	-.301	.577	.272	1	.602	.740
	Age(3)	.680	.589	1.333	1	.248	1.973
	Gender(1)	-.329	.211	2.442	1	.118	.719
	Occupation			44.877	5	.000	
	Occupation(1)	-1.214	.389	9.746	1	.002	.297
	Occupation(2)	.928	.341	7.421	1	.006	2.529
	Occupation(3)	.579	.293	3.914	1	.048	1.784
	Occupation(4)	-.357	.341	1.100	1	.294	.700
	Occupation(5)	-.179	.405	.195	1	.659	.836
	Education			20.613	3	.000	
	Education(1)	1.704	.436	15.296	1	.000	5.497
	Education(2)	1.592	.397	16.118	1	.000	4.916
	Education(3)	1.038	.391	7.058	1	.008	2.822
	Constant	-2.904	.753	14.859	1	.000	.055

a Variable(s) entered on step 4: Education.

Table 3.27: Model Summary (for model at step 4)

Step	-2 Log likelihood
1	887.392(a)

a Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table 3.28: Omnibus Tests of Model Coefficients (for model at step 4)

		Chi-square	df	Sig.
Step 1	Step	23.585	3	.000
	Block	23.585	3	.000
	Model	195.396	12	.000

There is a reduction in deviance by 23.585, which is a significant reduction, due to inclusion of educational level to the model at Step 3. So, educational level, in the presence of those variables at Step 3 has significant importance to describe response variable.

Step 5: Inclusion of Marital status to the model at Step 4.

Table 3.29: Variables in the Equation (for model at step 5)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Age			2.553	3	.466	
	Age(1)	-.052	1.210	.002	1	.966	.949
	Age(2)	.344	.576	.356	1	.551	1.410
	Age(3)	.643	.583	1.216	1	.270	1.902
	Gender(1)	.041	.234	.031	1	.861	1.042
	Occupation			13.978	5	.016	
	Occupation(1)	-.671	.411	2.659	1	.103	.511
	Occupation(2)	.271	.373	.527	1	.468	1.311
	Occupation(3)	.407	.317	1.647	1	.199	1.502
	Occupation(4)	-.453	.363	1.556	1	.212	.636
	Occupation(5)	-.155	.418	.137	1	.711	.856
	Education			14.750	3	.002	
	Education(1)	1.390	.451	9.504	1	.002	4.014
	Education(2)	1.467	.404	13.207	1	.000	4.334
	Education(3)	.971	.396	6.007	1	.014	2.641
	Mar.status			79.203	3	.000	
	Mar.status(1)	-1.295	.338	14.726	1	.000	.274
	Mar.status(2)	-2.917	.370	62.012	1	.000	.054
	Mar.status(3)	-.871	.360	5.849	1	.016	.419
	Constant	-1.454	.806	3.249	1	.071	.234

a Variable(s) entered on step 5: Mar.status.

Table 3.30: Model Summary (for model at step 5)

Step	-2 Log likelihood
1	801.412(a)

a Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Table 3.31: Omnibus Tests of Model Coefficients (for model at step 5)

		Chi-square	df	Sig.
Step 1	Step	85.980	3	.000
	Block	85.980	3	.000
	Model	281.376	15	.000

From Table 3.31, it can be seen that inclusion of marital status to the model at Step 4 results in reduction of deviance by 85.980. This indicates that including marital status to the model, in presence of those variables at Step 4, has significant importance.

Step 6: Inclusion of religion to the model at Step 5

Table 3.32: Variables in the Equation (for model at step 6)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Age			2.563	3	.464	
	Age(1)	-.075	1.381	.003	1	.957	.928
	Age(2)	.567	.585	.941	1	.332	1.764
	Age(3)	.793	.589	1.812	1	.178	2.210
	Gender(1)	.134	.237	.321	1	.571	1.144
	Occupation			13.669	5	.018	
	Occupation(1)	-.683	.411	2.757	1	.097	.505
	Occupation(2)	.342	.378	.819	1	.366	1.407
	Occupation(3)	.360	.319	1.278	1	.258	1.434
	Occupation(4)	-.478	.365	1.710	1	.191	.620
	Occupation(5)	-.071	.425	.028	1	.867	.931
	Education			16.235	3	.001	
	Education(1)	1.496	.457	10.701	1	.001	4.463
	Education(2)	1.544	.409	14.245	1	.000	4.685
	Education(3)	1.011	.401	6.352	1	.012	2.747
	Mar.status			80.551	3	.000	
	Mar.status(1)	-1.367	.345	15.688	1	.000	.255
	Mar.status(2)	-3.035	.381	63.595	1	.000	.048
	Mar.status(3)	-.973	.369	6.952	1	.008	.378
	Religion			11.533	4	.021	
	Religion(1)	-.035	.867	.002	1	.968	.966
	Religion(2)	-.487	.870	.314	1	.576	.614
	Religion(3)	-1.274	.954	1.783	1	.182	.280
	Religion(4)	-.912	1.181	.596	1	.440	.402
	Constant	-1.382	1.188	1.353	1	.245	.251

a Variable(s) entered on step 6: Religion.

Table 3.33: Model Summary (for model at step 6)

Step	-2 Log likelihood
1	788.410(a)

a Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Table 3.34: Omnibus Tests of Model Coefficients (for model at step 6)

		Chi-square	df	Sig.
Step 1	Step	13.002	4	.011
	Block	13.002	4	.011
	Model	294.378	19	.000

These summary and test of model coefficients show that the resulting reduction in deviance (13.002), due to inclusion of religion to the model at Step 5 has no significant importance, at 1% level of significance, to explain the response variable. Therefore, religion is omitted (not included) from (in) the model due to its insignificance.

Step 7: Inclusion of Residence area to the model at Step 5

Table 3.35: Variables in the Equation (for model at step 7)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Age			2.519	3	.472	
	Age(1)	.068	1.220	.003	1	.956	1.070
	Age(2)	.379	.578	.431	1	.512	1.461
	Age(3)	.672	.584	1.324	1	.250	1.958
	Gender(1)	.054	.234	.053	1	.817	1.056
	Occupation			13.431	5	.020	
	Occupation(1)	-.694	.412	2.841	1	.092	.499
	Occupation(2)	.263	.373	.495	1	.482	1.300
	Occupation(3)	.377	.319	1.400	1	.237	1.459
	Occupation(4)	-.450	.363	1.540	1	.215	.638
	Occupation(5)	-.171	.419	.167	1	.683	.843
	Education			14.920	3	.002	
	Education(1)	1.407	.451	9.711	1	.002	4.084
	Education(2)	1.474	.404	13.342	1	.000	4.368
	Education(3)	.978	.396	6.085	1	.014	2.658
	Mar.status			79.533	3	.000	
	Mar.status(1)	-1.299	.338	14.803	1	.000	.273
	Mar.status(2)	-2.923	.371	62.211	1	.000	.054
	Mar.status(3)	-.863	.361	5.729	1	.017	.422
	Residece(1)	.208	.279	.552	1	.458	1.231
	Constant	-1.664	.855	3.791	1	.052	.189

a Variable(s) entered on step 1: Residece.

Table 3.36: Model Summary (for model at step 7)

Step	-2 Log likelihood
1	800.847(a)

a Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Table 3.37: Omnibus Tests of Model Coefficients (for model at step 7)

		Chi-square	df	Sig.
Step 1	Step	.565	1	.452
	Block	.565	1	.452
	Model	281.941	16	.000

This last table indicates that the reduction in deviance due to inclusion of residence area in the model at Step 5 is 0.565, which is not a significant change. This implies including this variable in the model containing variables at Step 5 has no importance with regard to explaining the response variable.

3.3.2 Test of Goodness of Fit of the Final Model (Overall fit of Model at Step5)

The hypothesis to be tested here is:

$\mathbf{H_0}$: The model fits the data.

$\mathbf{H_A}$: The model does not fit the data.

To carry out this test, deviance analysis cannot be used due to inaccuracy because of the limitation noted in Section 2.4.2.1.The appropriate test of goodness to the data is the Hosmer and Lemeshow's goodness of fit test. The SPSS output of this test for our final model is given in Table 3.38.

Table 3.38: Contingency Table for Hosmer and Lemeshow Test

		Result = negative		Result = positive		
		Observed	Expected	Observed	Expected	Total
Step 1	1	163	164.461	3	1.539	166
	2	129	128.807	2	2.193	131
	3	126	125.963	3	3.037	129
	4	145	147.049	6	3.951	151
	5	138	136.373	4	5.627	142
	6	149	147.581	7	8.419	156
	7	147	143.838	10	13.162	157
	8	121	125.471	30	25.529	151
	9	96	99.773	51	47.227	147
	10	69	63.684	62	67.316	131

The test (using the observed and expected values from Table 3.38 in formula (2.11)) gives the following result.

Table 3.39: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.329	8	.610

This shows that there is no sufficient evidence to reject the null hypothesis and it confirms that our model has a good fit.

Therefore, the best fit of the data is given as:

$$\log\left(\frac{p_i}{1-p_i}\right) = -1.454 + \beta_a^* A_i + \beta_g^* G_i + \beta_o^* O_i + \beta_e^* E_i + \beta_m^* M_i \qquad (3.1)$$

where, p_i = probability that the i^{th} individual will contract HIV, under given levels of the explanatory variables.

β_a^* = coefficient of the a^{th} category of age.

A_i = age of i^{th} individual.

β_g^* = coefficient of the g^{th} category of age.

G_i = gender of the i^{th} individual.

β_o^* = coefficient of the o^{th} category of occupation.

O_i = occupation of the i^{th} individual.

β_e^* = coefficient of the e^{th} category of educational level.

E_i = educational level of the i^{th} individual.

β_m^* = coefficient of the m^{th} category of marital status.

M_i = marital status of the i^{th} individual.

* All values of β^*'s are given in Table 3.29.

For instance, for a man in the age group 14-30, who is a secondary school student and single, the model becomes:

$$\log\left(\frac{p_i}{1-p_i}\right) = -1.454 + 0.344A_i + 0.041G_i - 0.671O_i + 0.971E_i - 2.917M_i \qquad (3.2)$$

Let the right hand side of (3.1) be μ. Then, from (3.1) we get:

$$\left(\frac{p_i}{1-p_i}\right) = e^{\mu} \qquad (3.3)$$

Then, using (3.3) we can compute the odds of any category relative to the reference category of the same variable. For a category having a logistic regression coefficient β, the odds ratio of that category relative to the reference category of the same variable (holding the effects of other explanatory variables constant) is given as:

$$\text{OR} = e^{\beta} = \exp(\beta) \qquad (3.4)$$

For explanatory variables having more than two categories, it is possible to compute odds ratio between two non reference categories. Let category 1 has a logistic regression coefficient β_1 and category 2 has coefficient β_2, then odds ratio of category 1 to that of category 2 is given as:

$$\text{OR} = \frac{odds(category\,1)}{odds(category\,2)} = e^{\beta_1 - \beta_2} = \exp(\beta_1 - \beta_2) \qquad (3.5)$$

Odds Ratio is useful to compare the level of exposure to HIV infection (HIV Risk) of one category to the other category under the same factor.

Moreover, it is also possible to compute the probability of one's getting infected with HIV from expression (3.2). That is:

$$p_i = \frac{e^{\mu}}{1 + e^{\mu}} \qquad (3.6)$$

Some numerical illustrations on how to make uses of (3.2) and (3.3) with their respective interpretations are given in Chapter 4.

Chapter Four

Conclusions and Recommendations

4.1 Interpretations and conclusions

The results obtained in Chapter 3 are good indicators for organizations, individuals, experts in different areas at different levels and the society at large engaged in prevention and control of HIV/AIDS. But all the results should be interpreted and given meanings in statistical language and using statistical concepts.

In Sections 3.1.1 through 3.1.7, the statistical tests of association indicated that the explanatory variables included in this study, except one's religion and residence area, are all associated with HIV infection. That is, HIV infection has association with one's age, gender, marital status, occupation and educational level. By this we mean that HIV infection rate is not the same (not evenly distributed) at all categories of those variables.

Statistical tests of association, conducted in this study, lead to the following conclusions.

- HIV infection spread is different in different age groups of people. Some biological, economic and demographic reasonings can be given to this association (by area experts).
- HIV infection has an association with gender of individuals. This study indicated that females are highly affected by the epidemic as compared with males. Social, economic and also biological barriers may have contributed their shares for this association. More reasons may be identified by different area experts.
- HIV infection rate also differs in different marital statuses. That is, the spread of the epidemic has association with one's marital status. One important finding of this study is that marriage cannot be considered as guarantee to not being infected. Married people are highly affected than singles.

- HIV infection has no association with one's residence area (urban or rural). This indicates that people living wherever are exposed almost equally to the virus.
- One's occupation is significantly associated with HIV infection. Drivers and soldiers, next to housewives are among the highly affected segments of the society.
- HIV infection has no statistically significant association with one's religion. Despite different religious groups, people are almost equally exposed to HIV infection. This may indicate that either none of the religious groups gives due attention to protect its members from HIV infection, or all religious communities act the same way against HIV infection. But, the latter reasoning is unlikely.
- Educational level of people has a significant influence on HIV infection. Illiterate people are the most affected group of people. But this does not mean that the highly educated people are not affected at all. About 2.7% of people with higher educational level attendance are infected with HIV.

In this study, statistical tests were also conducted to investigate whether any two explanatory variable have a joint effect on HIV infection. The results of these conditional tests bring us to the following conclusions:

- Given a person is HIV positive, the following pairs of explanatory variables (age, marital status), (age, religion), (age, educational level), (gender, residence area), (gender, religion), (gender, educational level), (marital status, residence area), (marital status, religion), (residence area, occupation), (residence area, religion), (residence area, educational level), (religion, occupation), and (religion, educational level) have disjoint (independent) effect on bringing HIV infection.

- All the remaining possible pairs of explanatory variables have associated (joint) effect on bringing HIV infection.

Interpretations (of results) and conclusions based on results obtained in Section 3.3 are the following (these results are from Table 3.29 for the best fit model obtained at Step 5 of that Section):

- The odds ratio of contracting HIV for age group 13 years or below to those of 50 years and older group (reference category) is 0.928. This implies that the odds of contracting HIV for the lowest age group is 0.928 times that of the upper most age group, showing that the lowest age group is at relatively lower risk of HIV infection as compared to the upper age group. But, the odds ratio of the age group 14-30 years to those of upper most age group is 1.410, implying that odds of contracting HIV at age group 14-30 years is 1.41 times that at upper most age group. Using (3.5), we can see that odds ratio of age group 31-49 year to that of 13 years or below age group is $e^{0.643-(-0.052)} = e^{0.695}$ =2.004, implying that odds of age group 31-49 years is 2.004 times that of age group 13 or less years. This last result may be a good indication for major mode of transmission of HIV to be sexual contacts. Because, sexual activity is much higher in age group 31-49 years than that of 13 or less years.
- The ratio of odds of contracting HIV for illiterate people relative to those who have higher educational level is 4.014, showing that the odds of contracting HIV for illiterate people is 4.014 times that for people at higher educational level.

- The odds ratio of females to males is 1.042. This indicates that the risk (odds) for females is 1.042 times that for males, indicating that females are at higher risk as compared to males.
- Similar inferences based on odds ratios can be drawn using the results of this study given in Table 3.29.
- The estimates of logit coefficients of the final model are all significant for all categories of educational level. This implies that educational level is the most important factor, as compared to others, to explain the response variable.
- Using (3.1) and (3.6), the probability of contracting HIV can be calculated for any person. For instance, consider a 34 years old married man who is a primary school graduate and currently working as a merchant. The logistic regression equation for such a person is given as:

$$\log\left(\frac{p}{1-p}\right) = -1.454 + 0.643(1) + 0.041(1) - 0.453(1) + 1.467(1) - 1.295(1)$$

From (3.6) we get, $p = \frac{e^{-1.142}}{1+e^{-1.142}} = 0.242$. Using similar procedure and results in Table 3.29, we can express one's chance of contracting HIV numerically.

4.2 Recommendations

As this is a statistical study and the work of a statistician is to investigate a statistically true result regarding some proposition and give appropriate interpretations of those results, much recommendations are expected from some area experts. Nonetheless, the following points are forwarded as recommendations:

- ✓ Test results in the first Section of Chapter 3 show that HIV infection has no significant association with one's religion and residence area. These indicate that all religions and residence areas are almost equally exposed to HIV infection. So, prevention efforts should put by all religious groups (by the religious leaders), and in all residence areas (urban and rural areas) equally. Most of the prevention activities, usually, are concentrated in urban areas. But this study shows that people in rural areas are also equally exposed to the epidemic, suggesting that the rural society needs equal attention as urban society. This study also shows that HIV infection has association with age, gender, marital status, occupation, and educational level. Therefore, more attention should be given in preventing people at higher risk categories of those variables.
- ✓ Some explanatory variables are seen to have joint (associated) effect on HIV infection. From this we recommend that, for efficient prevention, resources should be allocated to control one of the jointly acting (compounding) variables. Because, if we control one of the associated variable, the joint effect will be controlled. For instance, gender and occupation have joint influence on HIV infection. Then, rather than investing resources to control two variables separately, it is advisable to concentrate efforts and recourses on one of the variables, which will have control effect on the other one.

- ✓ The risk of HIV on illiterates is 4.014 times that on the educated people. Therefore, it is recommendable to educate all segments of the society to reduce the risk of HIV epidemic.
- ✓ More recommendations and applications can be obtained from different area experts. Such recommendations should be accompanied with appropriate interpretations of the results obtained here.

Some Overall Limitations of the Study

1. Due to cultural values and trends, respondents usually fail to report exact age, exact educational level, exact occupation and the like. That is, data collected on such explanatory variables may lack quality to some extent. So, results in this study should be consumed taking those barriers into consideration.
2. The most recommended statistical package to handle and analyze categorical variables, especially to build logistic regression, is the GLIM package. But, due to budget constraint, there was no access to get and use this package. So, all the analyses in this study are carried out by using the SPSS package, which may lack some important features to analyze such data. Therefore, the results of this study should not be taken as 100% accurate.
3. Due to time constraint and width of the study, interactions of the explanatory variables are not given due attention and analyzed in logistic regression development. Of course inclusion of the effects of interaction terms may or may not improve the fit of model than the one we have now. However, the omission of such terms from the model is due to the mentioned constraint.

Appendix:

Table A1: **Age * Result Crosstabulation**

			Result		
			negative	positive	Total
Age	13 years and bellow	Count	20	1	21
		Expected Count	18.4	2.6	21.0
		% within Age	95.2%	4.8%	100.0%
		% within Result	1.6%	.6%	1.4%
		% of Total	1.4%	.1%	1.4%
	14- 30 years	Count	1086	116	1202
		Expected Count	1055.6	146.4	1202.0
		% within Age	90.3%	9.7%	100.0%
		% within Result	84.6%	65.2%	82.3%
		% of Total	74.3%	7.9%	82.3%
	31-49 years	Count	165	56	221
		Expected Count	194.1	26.9	221.0
		% within Age	74.7%	25.3%	100.0%
		% within Result	12.9%	31.5%	15.1%
		% of Total	11.3%	3.8%	15.1%
	50 years and above	Count	12	5	17
		Expected Count	14.9	2.1	17.0
		% within Age	70.6%	29.4%	100.0%
		% within Result	.9%	2.8%	1.2%
		% of Total	.8%	.3%	1.2%
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0
		% within Age	87.8%	12.2%	100.0%
		% within Result	100.0%	100.0%	100.0%
		% of Total	87.8%	12.2%	100.0%

Table A2 : **Gender * Result Crosstabulation**

			Result		Total
			negative	positive	
Gender	male	Count	679	65	744
		Expected Count	653.4	90.6	744.0
		% within Gender	91.3%	8.7%	100.0%
		% within Result	52.9%	36.5%	50.9%
		% of Total	46.5%	4.4%	50.9%
	female	Count	604	113	717
		Expected Count	629.6	87.4	717.0
		% within Gender	84.2%	15.8%	100.0%
		% within Result	47.1%	63.5%	49.1%
		% of Total	41.3%	7.7%	49.1%
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0
		% within Gender	87.8%	12.2%	100.0%
		% within Result	100.0%	100.0%	100.0%
		% of Total	87.8%	12.2%	100.0%

Table A3: **Marital status * Result Crosstabulation**

			Result		Total
			negative	positive	
Mar.status	married	Count	200	65	265
		Expected Count	232.7	32.3	265.0
		% within Mar.status	75.5%	24.5%	100.0%
		% within Result	15.6%	36.5%	18.1%
		% of Total	13.7%	4.4%	18.1%
	single(unmarried)	Count	990	39	1029
		Expected Count	903.6	125.4	1029.0
		% within Mar.status	96.2%	3.8%	100.0%
		% within Result	77.2%	21.9%	70.4%
		% of Total	67.8%	2.7%	70.4%
	divorced(separated)	Count	72	44	116
		Expected Count	101.9	14.1	116.0
		% within Mar.status	62.1%	37.9%	100.0%
		% within Result	5.6%	24.7%	7.9%

		% of Total	4.9%	3.0%	7.9%
		Std. Residual	-3.0	7.9	
		Adjusted Residual	-8.8	8.8	
	widowed	Count	21	30	51
		Expected Count	44.8	6.2	51.0
		% within Mar.status	41.2%	58.8%	100.0%
		% within Result	1.6%	16.9%	3.5%
		% of Total	1.4%	2.1%	3.5%
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0
		% within Mar.status	87.8%	12.2%	100.0%
		% within Result	100.0%	100.0%	100.0%
		% of Total	87.8%	12.2%	100.0%

Table A4 : **Residence area * Result Crosstabulation**

			Result		
			negative	positive	Total
Residence	urban	Count	1120	156	1276
		Expected Count	1120.5	155.5	1276.0
		% within Residence	87.8%	12.2%	100.0%
		% within Result	87.3%	87.6%	87.3%
		% of Total	76.7%	10.7%	87.3%
	Rural	Count	163	22	185
		Expected Count	162.5	22.5	185.0
		% within Residence	88.1%	11.9%	100.0%
		% within Result	12.7%	12.4%	12.7%
		% of Total	11.2%	1.5%	12.7%
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0
		% within Residence	87.8%	12.2%	100.0%
		% within Result	100.0%	100.0%	100.0%
		% of Total	87.8%	12.2%	100.0%

Table A5: **Occupation * Result Crosstabulation**

			Result		
			negative	positive	Total
Occupation	student	Count	478	13	491
		Expected Count	431.2	59.8	491.0
		% within Occupation	97.4%	2.6%	100.0%
		% within Result	37.3%	7.3%	33.6%
		% of Total	32.7%	.9%	33.6%
	housewife	Count	82	49	131
		Expected Count	115.0	16.0	131.0
		% within Occupation	62.6%	37.4%	100.0%
		% within Result	6.4%	27.5%	9.0%
		% of Total	5.6%	3.4%	9.0%
	unskilled lab&unemployed	Count	203	61	264
		Expected Count	231.8	32.2	264.0
		% within Occupation	76.9%	23.1%	100.0%
		% within Result	15.8%	34.3%	18.1%
		% of Total	13.9%	4.2%	18.1%
	merchant	Count	168	20	188
		Expected Count	165.1	22.9	188.0
		% within Occupation	89.4%	10.6%	100.0%
		% within Result	13.1%	11.2%	12.9%
		% of Total	11.5%	1.4%	12.9%
	Gov&NGO employee	Count	194	13	207
		Expected Count	181.8	25.2	207.0
		% within Occupation	93.7%	6.3%	100.0%
		% within Result	15.1%	7.3%	14.2%
		% of Total	13.3%	.9%	14.2%
	Others (includes police, military, drivers...)	Count	158	22	180
		Expected Count	158.1	21.9	180.0
		% within Occupation	87.8%	12.2%	100.0%
		% within Result	12.3%	12.4%	12.3%
		% of Total	10.8%	1.5%	12.3%
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0
		% within Occupation	87.8%	12.2%	100.0%
		% within Result	100.0%	100.0%	100.0%
		% of Total	87.8%	12.2%	100.0%

Table A6: **Religion * Result Crosstabulation**

			Result		Total
			negative	positive	
Religion	Orthodox Christian	Count	666	107	773
		Expected Count	678.8	94.2	773.0
		% within Religion	86.2%	13.8%	100.0%
		% within Result	51.9%	60.1%	52.9%
		% of Total	45.6%	7.3%	52.9%
	Protestant	Count	440	59	499
		Expected Count	438.2	60.8	499.0
		% within Religion	88.2%	11.8%	100.0%
		% within Result	34.3%	33.1%	34.2%
		% of Total	30.1%	4.0%	34.2%
	Muslim	Count	134	7	141
		Expected Count	123.8	17.2	141.0
		% within Religion	95.0%	5.0%	100.0%
		% within Result	10.4%	3.9%	9.7%
		% of Total	9.2%	.5%	9.7%
	Catholic	Count	20	2	22
		Expected Count	19.3	2.7	22.0
		% within Religion	90.9%	9.1%	100.0%
		% within Result	1.6%	1.1%	1.5%
		% of Total	1.4%	.1%	1.5%
	Others(including babies)	Count	23	3	26
		Expected Count	22.8	3.2	26.0
		% within Religion	88.5%	11.5%	100.0%
		% within Result	1.8%	1.7%	1.8%
		% of Total	1.6%	.2%	1.8%
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0
		% within Religion	87.8%	12.2%	100.0%
		% within Result	100.0%	100.0%	100.0%
		% of Total	87.8%	12.2%	100.0%

Table A7: **Educational level * Result Crosstabulation**

			Result		Total
			negative	positive	
Education	Illiterate	Count	111	42	153
		Expected Count	134.4	18.6	153.0
		% within Education	72.5%	27.5%	100.0%
		% within Result	8.7%	23.6%	10.5%
		% of Total	7.6%	2.9%	10.5%
	primary(KG-8)	Count	324	79	403
		Expected Count	353.9	49.1	403.0
		% within Education	80.4%	19.6%	100.0%
		% within Result	25.3%	44.4%	27.6%
		% of Total	22.2%	5.4%	27.6%
	Secondary(9-12)	Count	486	47	533
		Expected Count	468.1	64.9	533.0
		% within Education	91.2%	8.8%	100.0%
		% within Result	37.9%	26.4%	36.5%
		% of Total	33.3%	3.2%	36.5%
	Higher(10+1 &above)	Count	362	10	372
		Expected Count	326.7	45.3	372.0
		% within Education	97.3%	2.7%	100.0%
		% within Result	28.2%	5.6%	25.5%
		% of Total	24.8%	.7%	25.5%
Total		Count	1283	178	1461
		Expected Count	1283.0	178.0	1461.0
		% within Education	87.8%	12.2%	100.0%
		% within Result	100.0%	100.0%	100.0%
		% of Total	87.8%	12.2%	100.0%

Table A8: Source VCTs and sample size taken from each

Source VCT				
Bethzatha	Awassa Health Center	Family Planning	OSSA	Youth Center
420	335	196	278	232

References:

AIDS InfoNet, Fact Sheet No. 61(2004). **Older People and HIV.**

AIDS InfoNet, Fact Sheet No. 616 (2005). **Older People and HIV.**

Akukwe C. (1999). **HIV/AIDS in African children**: *a major calamity that deserves urgent global action.* Journal of HIV/AIDS prevention and education for adolescents & Children *Vol.* 3, *No. 3, pp.5-24.*

Brookmeyer, R. and M.H. Gail (1986). **A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic.** *Journal of American Statistical Association. Vol. 83, pp. 301-308.*

Caldwell, J. C. (Mar., 2000). **Rethinking the African AIDS Epidemic**: *Population and Development Review, Vol. 26, No. 1, pp. 117-135.*

Caldwell, J. C. (1997). **The Impact of the African AIDS Epidemic**. *Health Transition Review, No. 7, Supplement 2, pp.1-19.*

Caldwell, J. C. and J.K. Anafri (1997). 'Main' girlfriends, girlfriends, marriage, And money: *The social context of HIV risk behavior in Sub-Saharan Africa.*

Carael, M. (1997). **Urban-rural differentials in HIV/STDs and sexual behavior**: *Herdt G., Sexual cultures and Migration in the Era of AIDS: Anthropological and Demographic Perspectives; Clarendon Press, Oxford.*

Central Statistical Authority (CSA), Ethiopia (1994). **The 1994 Housing and Population Census.** *Result for SNNPR.*

Christensen, R., (1997). **Log-Linear Models and Logistic Regression** (2nd Edition): *Springer, New York.*

Clark, S. (2004) **Early Marriage and HIV Risk in Sub-Saharan Africa.**

Collett, D. (1991). **Modeling Binary Data.** *Chapman & Hall, London. New York. Tokyo. Melbourne. Madras.*

Cox, D.R., R.M. Anderson & H.C. Hiller (1989). **Epidemiological and statistical aspects of the AIDS epidemic**. *Phil. Trans. S. Soc. London. (B) 325: 3718.*

Day, N. E, S.M. Gore, M.A. McGee, M.South (1986). **Prediction of AIDS Epidemic in the U.K.**: *The use of the back projection method. Phil Trans R Soc London 1986; 325:123-34.*

Dejene Getahun (2005). **Youth Sexual Behaviour and the Risk of HIV Infection in Urban Ethiopia:** *The Case of Awassa City. Unpublished M. A. Thesis, Addis Ababa University, pp. 40-51*

Dyson T. (2003). **HIV/AIDS and Urbanization**: *Population and Development Review Vol. 29, No. 3, pp. 427-442.*

Family Guidance Association of Ethiopia (FGAE) (1998). **A Base line survey on Knowledge, Attitude and Practice of Sexuality and Reproductive Health among Jimma Youth:** *FGAE, Research and Evaluation Unit.*

FAO (2001). **THE IMPACT OF HIV/AIDS ON FOOD SECURITY**. *Available at web site: www.fao.org/docrep/meeting/003/Y0310E.htm*

Fienberg, S.E. (1979). **The use of chi-squared statistics for categorical data problem**. *Journal of the Royal Statistical Society, Series B, 41, 54-64.*

Gajjar, Y., C. Mehta, N. Patel, and P. Senchaudhuri (1998). **Statistical Package for Exact Nonparametric Inference (StatXact).** *Cytel Software, user manual.*

Glynn, J.R., M. Carael, A. Buve, R.M. Musonda, and M. Kahindo: Study Group on the Heterogeneity of HIV Epidemics in African Cities (2003). **HIV risks in relation to Marriage in areas with high prevalence of HIV infection**

Hosmer, D. W. and S. Lemeshow (2000) .**Applied logistic regression** (2nd Edition)**:** *New York: Wiley.*

Jordan-Harder, B., L. Maboko, D. Mmbando, G. Riedner, E. Nägele, J. Harder, V. Küchen, A. Kilian, R. Korte, F. Sonnenburg (2004). **Thirteen years HIV-1 sentinel surveillance and indicators for behavioural change suggest impact of programme activities in south-west Tanzania.**

Larntz, K. (1978). **Small sample comparisons of exact levels for chi-square goodness-of-fit statistics.** *Journal of the American Statistical Association, 73, 253-263.*

May, R. M., and R. M. Anderson (1987). **The Statistical Analysis and Mathematical Modeling of AIDS.**

Meehan, A. et al., (2004). **Prevalence and risk factors for HIV in Zimbabwean and South African women**. *XV International AIDS Conference pp. 11-16.*

Ministry of Health, Ethiopia (MOH) (2004). **AIDS in Ethiopia 5th edition**: *Disease Prevention and control department, MOH. Addis Ababa*

Minstry of Health (2002). **AIDS in Ethiopia, 4th edition.** *Addis Ababa, Disease Prevention and Control Department, MOH.*

National Intelligence Council, USA (2002). **The Next Wave of HIV/AIDS**: *Nigeria, Ethiopia, Russia, India, and China.*

O'Brien, W.A., P.M. Hartigan, D. Martin, J. Esinhart, A. Hill, S. Benoit, M. Rubin, M.S. Simberkoff and J. D. Hamilton (1996). **Changes in plasma HIV-1 RNA and CD4+ lymphocyte counts and the risk of progression to AIDS.** *Veterans Affairs Cooperative Study Group on AIDS.*

Population Reference Bureau (PRB) (2005). 2005 Population of the world Data Sheet. *Available at web site www. prb. Org*

The World Bank (2004). **Integrating Gender Issues in selected HIV/AIDS projects in the African Region.** *A Base line Assessment. Washington D.C., Africa Region Human Development working paper series No. 67.*

UNAIDS (1999). ***Listen, Learn, Live! World AIDS Campaign with Children and Young People:*** *Facts and Figures.* Geneva.

UNAIDS (2000). **Report on the Global HIV/AIDS Epidemic**. *Geneva*

UNAIDS (2002). **Report on the Global HIV/AIDS Epidemic**. *Geneva.*

UNAIDS (2004). **Report *on the global AIDS epidemic.*** *Geneva.*

UNAIDS (2005) **Intensifying HIV Prevention Policy Position Paper.** *Geneva*

UNAIDS (2005).**AIDS in Africa**: *Three Scenarios to 2025, pp. 27-60.*

UNAIDS/WHO (2005). **Progress on global access to HIV antiretroviral therapy:** *An update on '3 by 5'. UNAIDS/WHO. Geneva.*

UNICEF (2004). **Girls, HIV/AIDS and Education**. *New York.* Available at http://www.unicef.org/publications/index_25047.html

WHO (2003). **Integrating Gender into HIV/AIDS Programmes.**

WHO (2006). **Gender and HIV/AIDS.** ***Available at web site:*** ***www.who.int/gender/hiv_aids/en/***

Yang, H., **X.** Li , **B.** Stanton , **H.** Liu , **N.** Wang , **X.** Fang, **D.** Lin , **X.** Chen **.(2005).** **Heterosexual transmission of HIV in China:** *A systematic review of behavioral studies in the past two decades. Sexually Transmitted Diseases, 32(5):270-280. May.*

Zhang, K.L. S.J. Ma, D. Y. Xia (2004). *Epidemiology of HIV and sexually transmitted infections in China. Sexual Health, 16:39-46.*

Declaration

This Thesis is my original work and the first of its kind that has not done and presented elsewhere for any degree award.

Alemtshehai Abate

This Thesis has been submitted for examination in my approval as advisor.

Prof. Eshetu Wencheko